UNIVERSIDAD NACIONAL
AUTÓNOMA DE MÉXICO

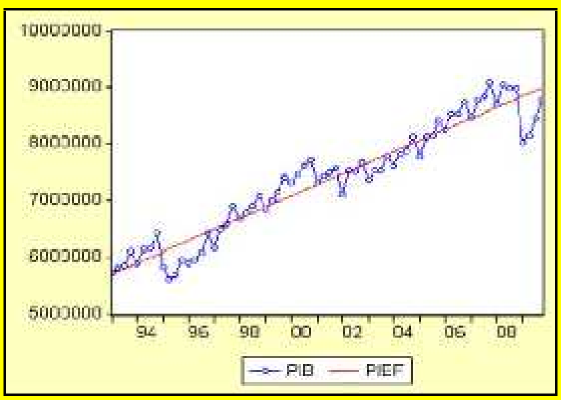
FACULTAD DE ECONOMÍA

Dirección General de Asuntos de Personal Académico

Programa de Apoyo a Proyectos para la Innovación y Mejoramiento de la Enseñanza

ECONOMETRÍA BÁSICA: CON LAS MODERNAS TÉCNICAS DE LA EDUCACIÓN DEL CONOCIMIENTO APLICADAS AL ANÁLISIS ECONÓMICO

RESPONSABLE: DR. GENARO SÁNCHEZ BARAJAS



Julio 2013

ECONOMETRÍA BÁSICA: CON LAS MODERNAS TÉCNICAS DE LA EDUCACIÓN DEL CONOCIMIENTO APLICADAS AL ANÁLISIS ECONÓMICO

PAPIME305811

ÍNDICE:

ECONOMETRÍA BÁSICA: CON LAS MODERNAS TÉCNICAS DE LA EDUCACIÓN DEL CONOCIMIENTO APLICADAS AL ANÁLISIS ECONÓMICO

Introducción

La aplicación de las innovaciones tecnológicas en diferentes campos del conocimiento en los últimos años, ha revolucionado el qué, cómo, porqué, dónde y cuándo hacer las cosas en cada uno de ellos. En este sentido destacan las NTIC: Nuevas tecnologías de la Información y las Comunicaciones, que se han dado a conocer con el nombre de "sociedad del conocimiento" y que ciertos estudiosos prefieren llamarla "educación del conocimiento". Cualesquiera que sea su denominación, para la elaboración de este libro, dado el titulo que tiene, considero que esta última acepción pone de manifiesto que su incidencia ya es muy grande en la formación de recursos humanos, dentro de la cual está comprendida la enseñanza de la econometría, disciplina que también ha cobrado mucha importancia porque sustancia *cuantitativamente* el análisis empírico con el que se coadyuva en la demostración de la tesis que sea de interés para el investigador. Las NTIC han venido a fortalecerlo al ayudar por medio de la *información* a la obtención y procesamiento de datos y, las *comunicaciones*, principalmente por medio del internet, también hacen más rápida la obtención de datos, pero además a la consulta de tablas estadísticas, de bibliografía, de metodología, de autores, etc. con relativas sencillez y rapidez.

Derivado de lo anterior es que la docencia de la econometría últimamente ha mejorado mucho debido a que los *conceptos teóricos se explican* con mejores ilustraciones y con mayor disposición de tiempo al ahorrar las NTIC el que antes se dedicaba a los planteamientos de los mismos. Ello ha favorecido la instrumentación práctica de los mismos, es decir, con las NTIC se hace con mayor propiedad tanto econometría pura como aplicada o empírica.

Es por esa razón que en los planes y programas de estudio universitario se hace énfasis en la incorporación de las NTIC como un medio eficaz para la realización de clases y prácticas de campo. En ese contexto, este libro responde a las exigencias planteadas ya que su objetivo fundamental es que los profesores además de mejorar la transmisión del conocimiento, ilustren su utilidad aplicándolo a casos reales de la economía y, que los estudiantes, se familiaricen con relativa facilidad con el tratamiento empírico del marco econométrico en el análisis de casos existentes. En esa forma los lectores estarán en mejores condiciones de adquirir por un lado la teoría, es decir, el sustento del análisis del *tema d*e su interés y, por el otro, de aplicar las técnicas econométricas en la solución de los problemas que éste tenga.

Es por lo anterior que el enfoque didáctico del libro se caracteriza por la exposición de la teoría econométrica, a la cual le sigue su aplicación a un caso concreto de la economía mexicana, en el que se van usando los conceptos teóricos apropiados para caracterizarlo, detectar su problemática y posible solución. Para ello se utiliza Eviews por ser un programa amigable que se puede adquirir fácilmente; con él se van indicando las acciones secuenciadas para la obtención de los datos que sustenten el análisis e interpretación de los resultados que se obtenga. En lo que se refiere al análisis e interpretación, se hace énfasis tanto en el estadístico como en el económico de los datos, porque de ello dependerá la toma de decisiones que se haga con respecto al fenómeno que se esté estudiando. En este sentido es que cada tema va acompañado de ejercicios resueltos y de reactivos con el objetivo de asentar el conocimiento de la teoría econométrica y de consolidar el aprendizaje en el manejo de los procedimientos de la información y de las comunicaciones.

Al respecto, conviene señalar que la oportunidad que nos brinda la UNAM a través de la Facultad de Economía de hacer un libro sobre econometría básica para estudiantes de licenciatura *que empiezan a tener conocimiento de esta disciplina en el quinto semestre, constituye un reto porque su nivel de estudios todavía*no les permite aplicar correctamente el método científico en la formulación de la teoría económica que se desea estudiar con el instrumental econométrico. Lo anterior implica que al no tenerla bien estructurada, posiblemente, ello incidirá en la especificación incompleta o inapropiada del modelo que, en turno, producirá estimadores de los parámetros que adolezcan de algunas de las propiedades estadísticas que deben tener para fundamentar con precisión la aplicación que se haga de ellos en los análisis de estructura, de predicción y de evaluación de políticas públicas económicas.

Agréguese a lo anterior que *al no estar seriadas las materias* cuyos temas constituyen la cimiente de la estructura teórica de la econometría, puede darse el caso de que los alumnos aún no hayan tomado un curso de matemáticas, lo cual afortunadamente ocurre con poca frecuencia; sin embargo, en el caso de la estadística, esta situación si se presenta más seguido.

En consecuencia, los profesores tenemos que acudir a nuestra *experiencia didáctica para suplir en la medida de lo posible los conocimientos básicos* que se requieren para enseñar la econometría en forma medianamente satisfactoria. Los enfoques, temas y resultados que se obtienen son distintos porque *algunos profesores* optan por hacer énfasis en aspectos básicos de las matemáticas y de la estadística a costa de no ver completo el programa de estudio exigido por la Facultad de Economía, además de que no usan la cibernética para facilitar los cálculos de los resultados a que se llegan con la aplicación de los métodos econométricos. Otros si ven todo el programa pero apresuradamente y con la sensación de que no se afianzaron debidamente los conocimientos econométricos entre los estudiantes, quienes en su mayoría actúan con el deseo de simplemente aprobar la materia, sin que, en consecuencia, les interese usar sus métodos en el futuro para hacer análisis e interpretación de las variables económicas de su interés. Por abundar sobre la docencia debemos decir que también hay profesores, que para dar cumplimiento al programa de estudios, acuden a tácticas como son el uso combinado de la exposición de la teoría econométrica en el salón tradicional de clases y, en su momento, del aula de cómputo para ilustrar la aplicación. En el otro extremo, existen profesores que todo el curso lo dan utilizando los hardwares y software disponibles en la Facultad de Economía; esta modalidad deja la sensación de que se enseñaron los métodos de la informática y de la econometría (medios) pero *que el análisis e interpretación de los resultados (fines) es escueto* y que quizás habría sido conveniente la realización de ejercicios <u>manualmente</u> sobre ciertos temas básicos.

Esta heterogeneidad posiblemente explique en cierta forma los altos índices de deserción y de reprobación que se observan semestralmente, y que sea unas de las razones por las cuales las autoridades académicas me hayan autorizado la elaboración de un libro que *pedagógica y didácticamente* pueda servir para la enseñanza apropiada de la econometría básica con el uso de las Nuevas Tecnologías de la Información y de las Comunicaciones, ***NTIC,*** además de que puede servir de consulta en la investigación y el ejercicio profesional de la economía.

Con esos antecedentes y deseando realizar un libro cuyos contenidos puedan coadyuvar a la mejor enseñanza y uso de la econometría, es que *invité a colaborar a ciertos profesores que tienen experiencia docente en la materia al igual que a algunos alumnos que ya aprobaron esta materia*, con el objetivo de que con nuestra visión de conjunto podamos formular los contenidos del citado libro que requiere el lector para superar los obstáculos que plantean las limitaciones, carencias y heterogeneidades antes mencionadas.

En este contexto es que hemos creído necesario empezar por establecer el marco teórico que tiene la econometría, ya que al ser éste un grupo de conceptos importantes que se utilizan para formular y argumentar ideas en pro de ciertos objetivos y metas, en este caso son los fundamentos de la teoría, las características y, alcance de la metodología econométrica, de manera congruente con el nivel de estudios que tiene un estudiante de economía, primordialmente y en general de las ciencias sociales. Con él tendremos una guía para la exposición formal, lógica y secuenciada del temario que exige la Facultad de Economía en el curso introductorio a la econometría. Se espera que con este procedimiento los estudiantes también conozcan las propiedades que tiene esta disciplina como método científico, al igual que la ilustración de su aplicación en el análisis e interpretación de los resultados a que se llega con su *metodología*, la cual se ilustra aplicándola en las variables económicas mexicanas de interés para el investigador o el profesionista que las usa para hacer economía aplicada.

Con este enfoque es que se empieza a desarrollar el guión temático del libro, usando para ello el instrumental que brinda el programa de computación EViews 5 y las NTIC, mostrando con ellos de manera secuenciada cómo se construye la teoría económica que se desea estudiar; enseguida, cómo se expresa matemáticamente y posteriormente, cómo se verifica estadísticamente.

Este *proceso docente* y en cierta forma de investigación nos lleva a la identificación del método científico apropiado, así como a su uso en la creación de conocimientos, en este caso sobre una teoría económica. Una vez formulada, se enseña a los alumnos a localizar las bases de los datos adecuados que permitan expresarla cuantitativamente. En virtud de que al contar con la información necesaria ya se está en condiciones de empezar a hacer econometría en el siguiente paso, dado que antes en el marco teórico ya se expusieron y definieron los conceptos que integran la econometría (entre los que destacan la definición de un modelo, de los diferentes tipos que existen para modelar la teoría económica, de sus características básicas, así como de los métodos que existen para su instrumentación correspondiente, etc.), se inicia la explicación e ilustración de los algoritmos de cada método con el que se arriba a resultados específicos. En este contexto se pone especial énfasis en el apoyo que brindan las NTIC así como en el análisis e interpretación de los resultados que se obtienen con ellas, dado que estas fases son muy importantes porque permiten cerciorarse de que se alcanzaron los objetivos cuando se hizo la planeación de la investigación en torno a cierta teoría de interés para el investigador.

Así, el guión temático comprometido comprende:

I.- Las NTIC, su desarrollo e influencia en la enseñanza-aprendizaje, en particular de la econometría;
I.1-Descripción de las características del Programa de Cómputo Eviews 5, de su simbología e instrucciones necesarias para usarlo en la computadora;
II.- Origen, evolución, definición y objetivos, alcance, limitaciones y perspectivas de la econometría.
III.- Aspectos fundamentales de la econometría.
IV.- El Modelo de regresión y correlación simple.MLS.
V.- El Modelo de regresión y correlación múltiple.MLM.
VI.- Las violaciones a los supuestos del método de MCO;
VI. 1.- Heteroscedasticidad.
VI.2.- Autocorrelación.
VI.3.- Multicolinealidad.
VII. Evaluación General de los Conocimientos Básicos de la Econometría.
VIII. Modelos con Cambio Estructural.
IX. Números Índice.
X. Series de Tiempo.
XI. Estacionariedad y Raíces Unitarias.
XII. Modelos Dinámicos: de Vectores Autorregresivos: VAR y VEC: Vectores de Errores Corregidos, Cointegracion y Pronóstico.

Conviene enfatizar que este libro tiene un objetivo principal: mostrar cómo mejora la enseñanza de la econometría básica usando las NTIC en las ciencias sociales, *en general*, y concretamente, en la licenciatura de economía, de finanzas y de los negocios, independientemente de que también se puede usar como libro de consulta en el mundo de las ciencias sociales, entre los investigadores, los estudiantes de posgrado y las personas (principalmente economistas) que ejercen esta profesión.

Por consiguiente para la consecución del objetivo toral, es decir, para la transmisión gradual y apropiada del conocimiento econométrico, aparte de usar las NTIC, se utiliza la **estrategia didáctica** de exponer en la primera parte lo que he dado en llamar la teoría (Capítulos I,II y III) y en la segunda (Capítulos IV, V, y VI) lo que denomino la práctica, es decir, su ilustración con los métodos econométricos fundamentales en forma *reiterada* periódicamente en lugares clave y sin abrumaral lector; esta exposición es acompañada de *reactivos* consistentes en ejercicios teóricos y numéricos debidamente secuenciados en cada uno de los capítulos antes descritos, todo ello con el fin de afianzar el conocimiento de la siguiente manera:

a) Al final de la exposición teórica debidamente sustentada con ejemplos ilustrativos numéricos y gráficos de cada uno de los capítulos, estos se complementan con la incorporación de cuestionarios con *preguntas* sobre los aspectos relevantes de cada punto del tema visto, las cuales se contestan primero, durante la sesión *programada*de trabajo en el seno del grupo y luego, en las *sesiones de repaso* relacionadas con el avance que tenga el tema en el curso presencial, así como con “tareas o investigaciones” que los alumnos debe de realizar en la biblioteca y/o en su casa ;

b) Se complementa con ejercicios adicionales resueltos y sin resolver con el fin de lograr aguzarla mente y la familiaridad de los alumnos con los *temas,* con el manejo de los *métodos* y con el análisis e interpretación de los *resultados* que se obtienen con ellos;

c) En ciertos reactivos se pide a los estudiantes/lectores que realicen investigaciones sobre determinadas peculiaridades de la metodología aplicada a casos concretos, cuyos resultados generalmente se comentan y comparan con mis puntos de vista vertidos en el marco teórico establecido en los capítulos I, II y III;

d) Derivada de la “capacitación” anterior, periódicamente, una vez terminadas las actividades de preparación antes descritas, el libro contiene *exámenes* de autoevaluación con el fin de que el lector confirme por si mismo que adquirió los conocimientos programados y que puede seguir adelante;

e) Al terminar los capítulos IV, V, y VI, con objeto de recapitular la teoría y su ilustración aplicada, se les pide a los alumnos/lectores que hagan una investigación documental sobre una teoría económica que pueda ser de interés suyo, por ejemplo, la teoría del empleo en México con el objetivo primordial de que vayan sopesando su presunta incorporación al mercado laboral en el futuro inmediato. Esta “práctica adicional” les ayuda para mejorar su compresión de la importancia de la econometría, motivo por el cual generalmente la hacen con entusiasmo;

Importancia del análisis e interpretación de los resultados. Este es de suma importancia porque es la razón ***primigenia del proceso de enseñanza-aprendizaje, dado que nuestro interés es formar analistas* y *no simplemente calculistas***. Así, una vez aplicada la metodología econométrica en forma sistematizada a los datos de la teoría económica que se desea verificar, ello se hace desde dos puntos de vista:

1.- Estadístico, mediante el cual *se explica* la calidad, bondad de ajuste o aproximación que tienen los valores de los estimadores a los valores reales de los parámetros, así como la fundamentación con rigor técnico de la relación que se observa que existe entre las variables establecidas en el marco de la teoría económica a verificar con estos métodos.

2.- El económico, en el que se constata si se verifica la teoría económica; y por consiguiente, que el analista está en condiciones de hacer análisis estructural, de predicción y de verificación de políticas económicas sustentadas en la teoría económica en estudio.

Enseñanza congruente con el nivel educativo de los alumnos.

Con este enfoque se modifica el contenido de los textos de econometría básica que suelen darle poca importancia a la metodología que debe usarse para la formulación de las teorías económicas; dado que los autores prefieren hacer énfasis en la metodología que se usa para probarlas y usar los datos en análisis de estructura, de predicción y de evaluación de políticas económicas, lo cual es muy cuestionable en virtud de que si no se privilegia la formulación adecuada de la teoría que se desea verificar, posteriormente surgen lo que se conoce como "problemas de especificación del modelo econométrico"; ésta situación conlleva a revisar su formulación, que podría reducirse o eliminarse si antes se le hubiera dado la importancia que tiene la formulación adecuada de la teoría económica.

En este contexto con el uso de las NTIC también hay un ***viraje de ciento ochenta grados*** en la relación entre maestro y alumnos, puesto que la enseñanza se vuelve más interesante al fluir con facilidad la metodología y su aplicación para obtener datos, para hacer su análisis e interpretación correspondiente, ya que con las NTIC se evitan los pasos inherentes a las demostraciones (que están en los cuatro apéndices: matemático, de la probabilidad, de la inferencia estadística e introductorio a las series de tiempo) del origen de los métodos (se trabaja con *axiomas* principalmente y lo estrictamente necesario con *teoremas*) que por supuesto son importantes cuando se hace *econometría pura o teórica*, pero que no tienen esa relevancia en situaciones en que los alumnos apenas empiezan a familiarizarse con la materia, ya que como antes se indicó, aun no son sólidos sus conocimientos sobre la formulación de las teorías económicas como tampoco de la probabilidad y estadística y mucho menos de la teoría sobre las series de tiempo. Así, dado que en el quinto semestre de la licenciatura en economía está programado este curso que en mi opinión debería darse en semestres más avanzados por las razones antes descritas, considero que la enseñanza debe ser congruente con el nivel educativo que tienen los estudiantes.

<u>Enseñanza caracterizada más por el enfoque axiomático que demostrativo de leyes y teoremas.</u>

Por consiguiente, cabe enfatizar que por el acervo de conocimientos referenciales que tienen los estudiantes de licenciatura para iniciar el aprendizaje de la econometría, la enseñanza de ésta se basa más en actividades de economía aplicada (econometría empírica) que en investigación teórica, es decir, somos usuarios del conocimiento no lo hacemos ni lo demostramos porque en el quinto semestre el propósito principal es transmitirlo (no crearlo) e ilustrarlo haciendo econometría empírica: investigación aplicada.

Así, con base en la definición de que un axioma es la *afirmación* de que una idea (conocimiento) es evidente por sí misma y por ende, que no necesita ningún tipo de comprobación para enunciarla, yo transmito con ese enfoque la teoría y los métodos econométricos exigidos por el Programa de Estudios de de Economía.

Obviamente en <u>ningún momento descuido el rigor</u>, la formalidad y origen matemático de la teoría y de los métodos econométricos, sólo que por las razones arriba apuntadas le dedico menos tiempo a los teoremas que se definen como *"afirmaciones que pueden demostrarse* de manera lógica a partir de un axioma o de otros teoremas antes demostrados." En otras palabras, los presento y explico lo estrictamente necesario e informo a los alumnos que quien esté interesado en conocerlos a fondo, que pueden consultar los apéndices matemático y estadísticos que aparecen al final de mi libro, además de que <u>podemos</u>

profundizar en ellos durante las asesorías personalizadas que doy en horarios específicos en la Facultad.

Este método de aprendizaje es congruente con la enseñanza de la econometría en la *Maestría, donde se profundiza sobre las demostraciones* de la metodología econométrica.

Enfoque de investigación real sustentada en la teoría aplicada al análisis de variables económicas mexicanas:

Lo antes dicho significa que al ilustrar el uso de la *metodología* en variables económicas mexicanas, en algunos casos, sobre todo en las *financiera*s, al *analizar los resultados* de su aplicación, no siempre se observará congruencia con la teoría económica formulada, en particular con los signos de los coeficientes y con la bondad de ajuste debido a la presencia de violaciones al modelo, etc. Estas inconsistencia se explican con base en el marco teórico econométrico, es decir, se dice porqué suceden, además de que se indica cómo se pueden superar para arribar a una ecuación de regresión " satisfactoria o limpia de irregularidades", misma que en ese punto ya sirve para hacer análisis de predicción, de estructura y para evaluar políticas públicas o privadas correctamente. Con este enfoque se prepara al lector para enfrentar imprevistos, a desarrollar su capacidad creativa y criterios para resolver situaciones en que por el tamaño de la muestra, la naturaleza y *distribución* de los datos observados de las series reales, los *resultados econométricos* no siempre corresponden al marco teórico establecido para la teoría económica formulada previamente..

Con esta orientación didáctica, también se atiende la sugerencia de Wooldridge (2009:12) de que se empiece enseñando econometría con datos de corte transversal porque se requiere de un marco teórico menor al que exige el manejo de los de series temporales. Debido a lo anterior es que la ilustración de los MLS: Modelo Lineal Simple y el MLG: Modelo Lineal General, empiezan con esa clase de datos.

En ese sentido, si lo que interesa es que la enseñanza de este curso introductorio se caracterice por la transmisión de los conocimientos básicos de la econometría y su aplicación correcta a la economía, es decir hacer econometría *aplicada,* entonces el *contenido de este libro* responde a ese objetivo, puesto que no descuida el rigor técnico al *incluir los cuatro apéndices antes mencionados, y simultáneamente, hace énfasis en la explicación de* las características del *método científico* que se usa para formular teorías económicas; además de que lo ilustro aplicándolo a las teorías que luego verifico con el instrumental econométrico contenido en el programa de cómputo Eviews 5. Al respecto, este último fue seleccionado porque está disponible prácticamente en todas partes, de manera que con él se pueden masificar la pedagogía y didáctica descrita en este libro.

Así, con el uso de *Eviews 5* (que es parte de la NTIC) para hacer los cálculos de los indicadores relevantes y simultáneamente usar el *internet* (*mejorar las* *comunicaciones en el aula*) para bajar tablas estadísticas, para obtener bases de datos, para aclarar ciertos conceptos o conocer algunos que surgen durante el proceso de enseñanza- aprendizaje en la clase presencial, ***el maestro deja de ser el único transmisor del conocimiento.***

Al respecto, es muy importante señalar que gracias a las NTIC y a la computadora, al *disponerse de más tiempo para pensar y reflexionar sobre un tema,* y puesto que ahora existe un cúmulo enorme de información disponible que permite profundizar sobre el mismo y que podría llevarse más tiempo del programado, se debe tener mucho cuidado en que su exposición sea congruente con el nivel de estudios que tengan los alumnos cuando van a tomar un curso de econometría. Así, con base en lo antes dicho en párrafos arriba, es aconsejable que cuando los profesores programan sus materias *antes vía encuestas se cercioren* de que los alumnos ya cursaron y aprobaron los de estadística y matemáticas mínimamente; si cursaron metodología de la investigación, mucho mejor. En este contexto lo más importante, al saber que las teorías económicas constituyen la razón de ser de la econometría, se sugiere que para la formulación de las mismas se comience estableciendo relaciones sencillas de variables económicas para que su complejidad no sea obstáculo para la enseñanza inicial de esta disciplina.

Uso de reactivos y ejercicios con NTIC.

Conviene mencionar que todos estos materiales han sido sometidos a una prueba personal conmigo mismo para cerciorarme de que los entiendo, que sé su alcance y puedo aplicarlos con propiedad en el ámbito económico, y que también fueron consensados con los alumnos en los grupos tanto de estadística como de econometría. Precisamente esas “terapias” dieron la pauta para determinar el tipo de reactivos académicos que conviene formular para asentar el conocimiento econométrico entre los alumnos, ya que su aplicación le permite al profesor afirmar el conocimiento y determinar el grado de entendimiento que ha alcanzado la materia entre los estudiantes que asisten a los *cursos presenciales*, donde cada uno de ellos en promedio es de 56 horas al semestre.

En esta perspectiva conviene decir que la nueva didáctica que aquí presentamos usando NTIC, también se fundamenta en el análisis y evaluación de los libros escritos por connotados autores en este campo dentro de los que destacan en el campo de la estadística los profesores norteamericanos Mason y Lind, así como en la econometría españoles como Carrascal et al, José Hernández y César Pérez, por señalar algunos de los más recientes.

Los reactivos que aquí se presentan se describen a continuación, mismos que son aplicados dentro de un ciclo de enseñanza-aprendizaje, que se inicia con el análisis de los temas que integran el *programa de estudios* exigido por la Facultad de Economía; enseguida los alumnos son familiarizados con los softwares que se usan en las aulas de cómputo donde además muestro cómo acceder al internet como medio de comunicación para hallar información complementaria al tema que se está exponiendo, para acceder a bases de datos de Banxico, INEGI, SE, al igual que a tablas estadísticas, a bibliografías especificas, etc. a la que le siguen los ejercicios manuales y electrónicos en que se aplican los métodos al análisis de *variables representativas de la economía mexicana*, como son el PIB, el empleo, la inversión, etc. Con ese referente es que luego planteo ejercicios en los que hago énfasis sobre ciertos aspectos que juzgo de importancia primordial, como son el análisis e interpretación de los resultados; en la siguiente etapa les solicito que hagan investigaciones sobre ciertos temas con objeto de que conozcan otros puntos de vista para que los comparen con los míos y así arriben a una visión clara de los mismos. El ciclo culmina con la presentación que hacen los alumnos de exámenes parciales cuyos resultados me permiten determinar si es necesario volver sobre ciertos temas con determinados tipo de datos, de métodos y de cierto número de ejercicios, así como considerar qué otros temas han sido asimilados satisfactoriamente y que, por consiguiente puedo continuar con el resto del programa de estudios.

Nuevo procedimiento para transmitir la teoría econométrica.

La situación que priva entre gran parte del alumnado de economía, que manifiesta cierta animadversión o poca preferencia por el uso de la econometría como método de análisis económico, ratifica la conveniencia de reflexionar sobre si o no se debe de modificar el enfoque tradicional de transmitir el conocimiento. Hacerlo no sólo los beneficiará a ellos en particular sino posiblemente a muchos otros del área de las ciencias sociales, ya que posiblemente los motivará a estudiar con entusiasmo al ver que avanzan con este enfoque que hace énfasis en hacer economía aplicada, es decir, econometría empírica , al observar que arriban a resultados rápidamente con las NTIC y que les queda tiempo para hacer análisis e interpretación con propiedad de los mismos, de manera que quizás además de aumentar la masa crítica, se reducirá la deserción y aumentará su asistencia presencial a los cursos que se impartan sobre esta disciplina.

En esta perspectiva es que ahora se ha juzgado conveniente empezar el libro destacando la importancia de las NTIC y las características de Eviews 5 antes de introducir el marco teórico, mismo que en su momento se expone (Capítulo III) con el fin de que desde el principio el lector disponga de los conocimientos necesarios para que posteriormente, cuando trabaje algún tema en especial, con una secuencia previamente establecida para no saturarlo de conocimientos sin sentido o desvinculados, se ubique porque ya tiene una referencia previa de su importancia y del uso en economía. En este contexto es que la transmisión del conocimiento será dinámica e interesante ya que para logarlo además de aplicar las técnicas de NTIC, de regular la exposición de los contenidos sustantivos de los temas, de la aplicación de reactivos para constatar la asimilación de la teoría, *despertará* en ellos la inquietud de inquirir sobre si realmente se requiere de una teoría para hacer análisis e interpretación económica- estadística. La respuesta a estas dudas las dará el marco teórico o conceptual que se propone, con el resultado adicional de que habrá aumentado su cultura econométrica y de que será un profesionista o investigador que sepa estudiar un fenómeno económico con rigor científico.

Sugerencias para los profesores de econometría de hoy.

Dado que existe una gran resistencia entre los alumnos a la práctica manual (sobre todo de las demostraciones matemáticas) *de métodos numéricos* utilizados para obtener resultados que les permitan analizarlos e interpretarlos en las investigaciones económicas que le sugerimos que realicen, sugiero que ahora se aproveche esta oportunidad para *reflexionar sobre si ha sido correcto el enfoque* con que se ha actuado en la transmisión de los conocimientos econométricos a los alumnos de la licenciatura. En este contexto surge la duda sobre si ello sea una de las causas por la que la reprueban o desertan muchos alumnos y, lo más grave, por la que la desechan como método de análisis económico.

Así, conviene que los profesores determinemos si debemos usar las NTIC como eje rector permanente de la enseñanza aprendizaje de la econometría, lo cual significaría exponer inclusive la *teoría* con este instrumental o, si sólo debemos usarlo para bajar de internet (comunicación*)* en el aula de cómputo la información necesaria para hacer en ella los *cálculos electrónicamente co*n algún programa de los que están instalados en las Facultades de las Universidades.

Amplitud temática del libro

Comprende los temas que exige el programa de estudios de la econometría en la licenciatura, es decir, la introducción al uso de los métodos de la econometría clásica y moderna.

Originalidad de esta obra

Difícil es tenerla sobre todo cuando muchos y reconocidos autores han precedido este libro. Sin embargo, creo que la originalidad se da en que está escrita en congruencia con el nivel de estudios que tienen los estudiantes, en el uso de las NTIC aaplicadas de manera integral a la economía, en su aplicación al análisis de variables mexicanas, *en la presentación y desarrollo de los temas con un énfasis eminentemente didáctico tanto para asimilar la teoría como para hacer econometría empírica*, así como por el *uso de reactivos* para afianzar el conocimiento entre los estudiantes, los académicos, los profesionistas y el público en general, que hacen economía aplicada.

Finalmente, deseo señalar que si bien, he contado con la valiosa ayuda de los profesores José Alberto Reyes de la Rosa, Gabriela Longoria Hernández, Joel Rojas Escudero, Fátima Liliana Nava Sesma y Roxana Sierra González, así como de las becarias Celtzin Beatriz Bonilla Sánchez, Lidia Bonilla Zarrazaga, María Luisa Jiménez López, Anaid Karina Limón Rivera, y Karina Valle Cruz, en la elaboración de este libro, es justo decir que yo soy el único responsable de todas los errores, incongruencias o defectos que se observen en los contenidos de sus temas.

I.- LAS NUEVAS TECNOLOGÍAS DE LA INFORMACIÓN Y DE LAS COMUNICACIONES, NTIC; SU DESARROLLO E INFLUENCIA EN LA ENSEÑANZA –APRENDIZAJE EN LA ECONOMETRÍA.

La innovación tecnológica en materia de información y de comunicación cuya aplicación se masificó durante los últimos treinta años empleándola en la creación y transportación de la información por medio de excelentes canales de comunicación, ha permitido la utilización de más y mejores datos para la expansión del conocimiento humano en prácticamente todas las ciencias, los cuales son manejados en las computadoras por medio de programas de computo amigables prácticamente desde la temprana edad del ser humano, situación que provoca cambios sustantivos en él por la rápida absorción del conocimiento que hace sobre las características del mundo en que se desarrolla, que al hacerlo, está en condiciones de educarse prematuramente y mejor, de especializarse y en general para aumentar su cultura. La oportunidad que le brinda la innovación tecnológica así descrita está sustentada en lo que se ha dado en llamar la sociedad o educación del conocimiento, misma que he usado intensamente en mis prácticas docentes desde hace diez años (La Facultad de Economía me ha publicado libros de estadística y econometría con este enfoque); con ella he revolucionado mis métodos y programas pedagógicos y de investigación tanto en los niveles educativos de licenciatura como de doctorado.

I.1.-La triada que fundamenta la enseñanza usando NTIC.

Considero que si la educación del conocimiento hace posible que el ser humano (headware) aplique la tecnología (hardware) para captar mediante el internet y manipular la información con programas de cómputo (software) para transformarla en producto, es indudable que brinda una opción para la mejor transmisión del conocimiento a los estudiantes, quienes además de así adquirir una sólida formación teórica, desarrollan con celeridad su capacidad creativa para ser profesionistas e investigadores competitivos al contar, en el caso de los economistas, con instrumentos que los auxilian para hacer análisis e interpretaciones apropiadas de los fenómenos económicos que suelen estudiar.

Para que ellos puedan generar ese producto, que no es más que la aplicación del buen acervo Adquirido y de su hábil instrumentación técnica en la solución de los problemas económicos que aquejan a la sociedad, requieren de cambios radicales no sólo en los contenidos de los programas de estudio, también en los sistemas de enseñanza aprendizaje.

I.2.-Nueva pedagogía.

Si por pedagogía se entiende el conjunto de acciones mediante las cuales el profesor educa a las personas, principalmente en los aspectos psicológico, físico e intelectual, que, en el caso de los estudiantes de la UNAM, la educación se ejerce como una actividad social que se debe caracterizar por la transmisión efectiva de los conocimientos de manera masiva. Podemos decir que en ella se conjugan y hacen congruentes las necesidades personales con las del grupo en que participan.

En este contexto destaca la pedagogía activa en que el binomio maestro- alumno tienen un rol fundamental; el primero como conductor, motivador y *usuario de las NTIC* para mejorar la enseñanza y, el segundo, como *receptor crítico responsable* de la calidad de su aprendizaje. En esta perspectiva conviene agregar que este libro por medio de sus contenidos coadyuva no sólo con la pedagogía activa sino también con la programada que es el fundamento para ir transmitiendo gradualmente el acervo de conocimientos sin producir saturación o cansancio entre los alumnos.

Ello conlleva al compromiso de elaborar un nuevo texto de econometría, que sea diferentes a los vigentes, cuya *obsolescencia* se expresa no sólo por la ausencia de un enfoque pedagógico acorde con el perfil escolar de los alumnos de la Facultad de Economía y de las ciencias sociales en general; además por la ausencia de bases de datos (sobre todo de la economía mexicana) y de softwares que faciliten la transmisión de sus contenidos hacia una generación nacida en la era de la electrónica y por consiguiente ávida de libros de texto caracterizados por el uso de la computadora, del internet y de programas de cómputo. Lo anterior, conduce al diseño de una nueva pedagogía, cuya connotación es la de *enseñar a aprender* dentro de la sociedad del conocimiento, pero no sólo a los alumnos, también al profesor quien ahora debe tener la humildad de aceptar que debe aprender todos los días a conocer el potencial de estos medios para enseñar adecuadamente a sus discípulos.

Así, quiérase o no ha surgido una nueva pedagogía, la cual tiene como referente básico la informática y borda en torno al uso de las NTIC, apuntaladas por la tecnología del internet que es el vehículo que ha hecho posible el surgimiento, expansión y rápida aplicación del conocimiento económico. De ahí que sea conveniente abrevar en ésta para determinar los nuevos espacios y circunstancias en que se debe educar sobre la ciencia económica en la UNAM.

Esta situación ahora induce a pensar *cómo se debe enseñar a aprender* y con qué libros se debe hacer para evitar el rezago de la UNAM con respecto a otras instituciones que enseñan economía en el país y en el extranjero. En efecto si en el aula el profesor era el principal emisor de conocimientos, ahora con las carreteras de la información se está en posibilidad de modificar o implementar nuevos programas educativos, cuyo sustento para el profesor deben ser libros que además de contener los conocimientos básicos sobre econometría, debe tener un claro sustento en la cibernética. Vistos así los nuevos libros, su alcance es muy grande porque deben hacer posible el uso en el aula de clase del internet, que tiene la capacidad de transportar palabras, archivos, imágenes, gráficas y así establecer una relación educativa entre tutores y alumnos sin más limitación que la capacidad de los servidores utilizados.

En este contexto es que diremos que los nuevos libros de econometría deben permitir a profesores y alumnos acceder conjuntamente a las bibliotecas virtuales, a diccionarios especializados, a bases de datos y a una amplia gama de softwares especializados que de manera enunciativa pero no limitativa, se pueden mencionar al Office, Spss, Eviews, Stata, etc.

I.3.-Sugerencias de métodos didácticos para la enseñanza de la econometría a los alumnos:

Si la <u>didáctica es una disciplina científica- pedagógica</u>, entonces podemos decir que es un área de la pedagogía que se encarga de aplicar los sistemas y métodos de enseñanza contenidos en ésta última. Por consiguiente si utiliza las *diversas técnicas y formas de enseñanza* para mejorar la transmisión de los conocimientos econométricos, en nuestro caso procuramos *adaptarlas* a las necesidades de los estudiantes de economía, los negocios, las finanzas y de las ciencias sociales en general.

Así, dado que la función principal de un docente es la transmisión adecuada del conocimiento, es conveniente decir que para lograrlo yo uso y sugiero *cuatro* *métodos*, los cuales son los siguientes: el del diálogo socrático combinado con el sistémico, el holístico y de Montessori. Con el del *diálogo,* con los alumnos inquiero, busco, hallo, reflexiono, evalúo y digiero el conocimiento. Esta actividad la complemento usando el método sistémico para descubrir y reconocer en los alumnos, sus aptitutes, afinidades y para brindarles mi orientación para poder cursar exitosamente la materia de econometría. Con el método holístico doy sustento técnico a la exposición de los conceptos econometricos ya que procedo explicándolos en forma rigurosa y sistemática, desde su fundamentación en los axiomas hasta los teoremas y leyes en que se sustentan. Finalmente, con el método Montessori procuro que aprendan con actividades lúdicas para desestresarlos después de intensas sesiones de trabajo de concentración y abstracción absorbiendo la teoría econométrica ya que sugiero "aprender haciendo con acciones lúdicas que, inclusive, favorecen la autoeducación".

Luego entonces la educación generada con estos métodos determina que se considere al *aprendizaje como un fenómeno integrado a la persona entera*, que lo hace diferente del viejo enfoque mecanicista que asigna a la naturaleza humana el carácter dual cuerpo- mente.

En ese tenor es que **para modelar el sistema de enseñanza adecuado** de econometría acostumbro recabar al inicio del semestre escolar entre los alumnos su opinión sobre:

a) ¿Cuáles son sus expectativas del curso? ¿Por qué lo tomaron conmigo?

b) ¿Cuántas horas a la semana le dedican a la televisión, a la computadora, Internet e inglés: leer, escribir, conversar y entenderlo?

c) ¿Qué importancia tiene para ellos este curso en relación con los demás que tomará?, etc.

Es mi opinión que en esa forma se trabaja en un sistema de enseñanza-aprendizaje que cobra vida, que es dinámico y que no es repetitivo *al pasar* de una enseñanza estática descriptiva basada en la mecanización a una enseñanza inquisitiva, cuestionadora que genera y usa fundamentos técnicos para dar vida a un modelo vivo de enseñanza-aprendizaje.

De esta manera la enseñanza de la econometría no es simplemente una transmisión del conocimiento, sino que además en su aplicación inmediata es útil para desarrollar *la capacidad de análisis e interpretación* que debe caracterizar a todo profesionista competitivo en la globalización actual de la economía.

Creo que con esta práctica docente se eliminan el aburrimiento, hacen atractiva la asistencia a clases y por consiguiente reducen la apatía, la deserción y desesperación típica del ser humano (de los alumnos) de cómo sobrevivir en el futuro inmediato, puesto que están adquiriendo conocimientos útiles para labrar su destino en el mercado, ya sea como empleados o como emprendedores de negocios.

Así pues, considero que al ***trabajar con un modelo sustentado en la información pedida a los alumnos y en el programa de estudios exigido por las autoridades docentes***, permite interactuar, preguntar y por consiguiente, dudar siempre en grupo y en consecuencia, contando con mi guía y orientación en lo que se refiere al rigor científico; de esa manera procuro dar respuestas coherentes, demostrables y actualizadas .Lo anterior además de romper el monólogo del profesor, produce sinergias entre los estudiantes, el profesor y generalizando, con el entorno universitario en que interactuamos.

Práctica docente.

1.- En el contexto metodológico antes descrito de la educación es que procuro crear escenarios de aprendizaje cercanos a la realidad en que los alumnos actuarán posteriormente como profesionistas.

2.- Expongo y evalúo el tema programado, es decir, destaco sus principales características e indico el alcance de su metodología, sus limitaciones y aplicaciones adecuadas a la economía; los alumnos no repiten ni memorizan, sino que asimilan, reflexionan, critican y analizan los conceptos que les he transmitido. Además busco la aplicación conveniente de los métodos estadísticos aprendidos.

3.- Para cerciorarme del grado de asimilación de la teoría adquirida por los alumnos, acostumbro hacer lo siguiente:

a).- Elaborar un cuestionario con preguntas clave o básicas sobre el tema que se ha visto; el cuestionario lo entrego a los alumnos para que lo contesten fuera de aulas y cuando lo regresan, juntos revisamos si fueron correctas las respuestas que escribieron, procediendo a hacer las correcciones pertinentes cuando no haya sido el caso; los estudiantes que contestaron bien la mayoría de los preguntas reciben una nota por su "participación en clase", como estímulo para su mejor desempeño en el futuro. Ello se acompaña con el trabajo en grupo que llevo a cabo a través de los reactivos que aplico para afianzar los conocimientos adquiridos.

b).- Enseguida les aplico exámenes parciales escritos u orales a los alumnos; para ello los preparo en la siguiente forma: les comunico con anticipación el tema sobre el que versará el examen, enseguida se les *deja ejercicios para resolver fuera del aula:* Dicha solución se revisa en el pleno de la clase: cuando los alumnos los resolvieron correctamente, les reconozco dicho mérito con puntuación sobre su participación en clase, cuando no resolvieron o hicieron mal los ejercicios, les corrijo e informo que el examen tratará sobre casos similares.

4.- Como puede observarse con este enfoque didáctico pretendo que el alumno razone, infiera, deduzca, medite, reflexione y no repita o memorice. En este sentido es que utilizo las tres aulas: la *tradicional* para presentarnos, conocernos e introducir y adminstrar los tiempos del curso, además para establecer los criterios de evaluación para calificar sus resultados al final del semestre. En el aula de *cómputo* hacemos con rapidez los cálculos, situación que permite disponer de más tiempo para hacer análisis e interpretación de los datos. El aula multimedia sirve para exponer y hacer énfasis en ciertos puntos cruciales o básicos del conocimiento.

Para exponer uso los tres siguientes ***canales de percepción***: auditivo, kinestésico y el visual, sobre todo en el aula multimedia y en la de cómputo.

5.- Una vez afianzado el conocimiento econométrico, principalmente los métodos, acostumbro pedir que los alumnos investiguen en otras fuentes bibliográficas con objeto de ampliar el conocimiento, de conocer otros enfoques y quizás, de entender mejor la aplicación del método en la ilustración del tema en estudio.

Así pues, no pretendo enseñar econometría per se, sino que le enseño al alumno a reflexionar sobre el alcance, uso y limitaciones de la econometría, en una atmósfera propicia para considerar esta disciplina como una de las principales herramientas para el análisis e interpretación de los fenómenos económicos que le toque estudiar en el ejercicio profesional. En este contexto es que termino haciendo una investigación económica mediante la cual indico cómo se plantea una *teoría económica*, cómo se expresa *matemáticamente* y cómo se verifica o rechaza estadísticamente. Algo similar les pide a los alumnos para que empiecen a desarrollar modelando una teoría económica, presentándola matemáticamente y probándola estadísticamente al *final del curso* de econometría.

I. 4.- Las características básicas del Programa EVIEWS 5, de su simbología y funciones necesarias para operarlo en computadora

Por la importancia que tiene el programa Eviews en este libro, concretamente en los métodos de enseñanza-aprendizaje de la econometría básica, he juzgado conveniente describir algunas de sus características, propiedades y alcance que se logra con su desarrollo aplicado, principalmente, al análisis de la economía mexicana.

Con ello se pretende que los lectores no encuentren obstáculos en su funcionamiento. Conviene decir que fue seleccionado porque desde el punto de vista didáctico es importante decir que contiene menús desplegables y ventanas que facilitan el seguimiento de lo que se está haciendo y por consiguiente, la asimilación clara y congruente de los conocimientos.

El programa Eviews es un paquete interactivo de econometría que proporciona un análisis sofisticado de datos, regresiones y herramientas para la predicción del comportamiento de valores futuros.

Gracias a Eviews se pueden desarrollar relaciones estadísticas de datos transversales y datos de panel de una forma sencilla, ya que muchas de las operaciones que se pueden implementar usando los menús, también se pueden introducir por la ventana de comandos.

Las operaciones más sencillas que admite e-views para la modificación de series de tiempo o del modelo en general son: suma (+), resta o diferencia (-), multiplicación o producto (*), división (/), potencia o función exponencial (^), raíz cuadrada (√). Obtención de logaritmo (log, ln) matriz, entre otras.

Es importante recordar que el programa e-views calcula primero las potencias y raíces cuadradas, después las multiplicaciones y divisiones; y posteriormente sumas y diferencias.

Las variables se escriben en la barra de comandos y el resultado se ve reflejado en la parte de mensaje. Supóngase que X, Y, Z son variables, sin embargo, para efectos visuales y prácticos, se implementaron números en cada uno de los ejemplos con Eviews. La forma de crear las operaciones básicas es la siguiente[1]:

Cuadro I.1

[1]Kozhan, Roman (2009) Financial Econometrics- with EViews, Ventus Publishing ApS.P.p 7 User´s Guide EViews 5.

Función	Descripción
View (Vista)	Muestra la visualización de la serie.
Procs (Procedimiento)	Activa procedimientos a aplicar a la serie.
Objets (Objetos)	Es el menú de almacenamiento y presentación del objeto.
Print (Imprimir)	Imprime el gráfico o la serie.
Name (Nombre)	Permite cambiar el nombre al objeto serie asignado.
Freeze (Congelar)	Genera una tabla con el contenido actual.
Edit +/- (Edición)	Activa y desactiva el modo de edición de datos.
Smpl +/- (Muestra)	Presenta los datos en periodos seleccionados o para el total del rango.
Label +/- (Etiqueta)	Muestra y oculta la etiqueta de la serie.
Wide +/- (Ancho)	Cambia la visualización de la tabla de vertical a horizontal .
InsDel (Insertar)	Inserta o borra objetos de la serie.
Title (Titulo)	Permite introducir un titulo al objeto tabla.
Sample (Muestra)	Cambia el periodo de muestra activo.
Genr (Generar)	Permite transformar la serie y generar una nueva variable.

Cuadro I.2

Funciones	**Descripción**
Genr	Genera directamente una operación entre variables.
Log(X)	Logaritmo natural.
exp(X) o @exp(X)	Función exponencial e^x.
Abs o @abs(X)	Valor absoluto ?X?.
Sqr o @Sqr(X)	Raíz cuadrada.
=sin(X)	Función Seno
=cos(X)	Función Coseno.
=asin(X)	Arco seno.
=acos(X)	Arco coseno.
=tan(X)	Función tengente.
rnd	Número aleatorio entre cero y uno.

Cuadro I.3

Funciones	Descripción
nrnd	Número aleatorio con media cero y varianza uno.
=obs(X)	Número de observaciones de X.
=se	Error estándar de la regresión.
=ssr	Suma de cuadrados de los residuos.
Cross(x,y)	Producto cruzado de x e y.
=cov(x,y)	Covarianza entre x e y.
=aic	Criterio de Información del Akaike
=coefcov(i,j)	Matrix de Covarianza de i,j
=coefs(i)	Valor del coeficiente “i” en la regresión.
=dw	El estadistico Durbin-Watson de la regresión.
=f	La F-estadística
=fprob	La probabilidad de la F-estadística

Cuadro I.4

Funciones	Descripción
=trend, @trend(n)	Variable ficticia de tendencia.
=cor(x,y[,s])	Covarianza entre X e Y.
=mean(x[,s])	Media para la serie X.
Sym	Crea una matriz simétrica.
=min(x[,s])	Mínimo valor de la serie X.
=max(x[,s])	Máximo de la serie X.
=stdev(x[,s])	Desviación estándar de la serie X.
=sum(x[,s])	Suma de la serie X.
=var(x[,s])	Varianza de la serie X.

Cuadro 1.5

Funciones	Descripción
=identity(i)	Crea una matriz identidad de dimensión “i”
=inverse(X)	Crea una nueva matriz que es la inversa de una matriz no singular X.
=rank(k)	Crea un nuevo escalar con rango.de la matriz “k”.
=trace(M)	Crea un nuevo escalar que contiene la traza de la matriz “M”
=seas(n)	Crea una variable ficticia.

Cuadro I.6

Distribución	Función
Chi-square	@cchisq(x,v), @dchisq(x,v), =qchisq(p,v), =rchisq(v)
F-distibución	@cfdist(x,v1,v2), @dfdist(x,v1,v2), =qfdist(p,v1,v2), =rfdist(v1,v1)
Normal(Gaussian)	@cnorm(x), @dnorm(x), =qnorm(p), =rnorm, nrnd
T-Student´s	@ctdist(x,v), @dtdist(x,v), =qtdist(p,v), =rtdist(v)

Cuadro I.7

Tipo de Función	Empieza con el Nombre
Distribución Acumulada (CDF)	@c
Densidad o probabilidad	@d
Inversa de CDF	@q
Generador del Número Aleatorio	@r

Cabe señalar que el programa Eviews 5 que aquí se utiliza es una nueva versión del iniciado para Windows de un grupo de herramientas cibernéticas diseñadas para hacer econometría empírica.

Con este programa se pueden hacer, como se verá más adelante, análisis de:

1.- Estructura;

2.- Predicción;

3.- Evaluación de políticas públicas y de la empresa;

El hardware que se requiere para instrumentar sus funciones es aquel en el que pueda instalarse el Windows. Debe tener una capacidad mínima de 4 megas de memoria ram; sería preferible que contará con 10 megas.

Con esta capacidad de procesamiento se está en condiciones de mostrar las técnicas econométricas de la econometría clásica y de la moderna, así como su tratamiento con las herramientas que brinda Eviews para su cálculo cibernéticamente.

II.-ORIGEN, EVOLUCIÓN, DEFINICIÓN, OBJETIVOS, ALCANCE, LIMITACIONES Y PERSPECTIVAS DE LA ECONOMETRÍA.

Al término del capítulo el lector estará en condiciones de entender:

- La definición actualizada y esencial de econometría;
- El origen del concepto;
- Su evolución de simple medición a relación relacionar variables económicas, así como de su enfoque tradicional o clásico a moderno o de series temporales;
- El nuevo alcance de esta disciplina como método ilustrativo de teorías económicas expresadas matemáticamente y constatadas estadísticamente;
- Las principales limitaciones que tiene;
- Su aplicación macro y micro económicamente, por orden de gobernanza ,tamaño de empresa , sectorial y regionalmente, cualitativa y cuantitativamente, así como en el cálculo y análisis de prospección económica;
- Estado del arte de la econometría;
- Su enseñanza con las Nuevas Tecnologías de la Información y de las Comunicaciones, NTIC,:
- Los diferentes tipos de datos que utiliza para hacer análisis estructural y de predicción;
- Distinguir las diferencias entre modelos deterministas y econométricos;
- La importancia de porqué usar variables aleatorias con su correspondiente "término de error" en la medición de la fuerza con que una(s) variable(s) independiente(s) impacta(n) las variaciones en los valores de la variable dependiente;
- El concepto de propensión marginal, su cálculo, análisis e interpretación correspondiente en economía.

II.1- Origen del vocablo.

En el año de 1926 el noruego Ragnar Friech acuñó el término ECONOMETRÍA cuya inspiración provino de la observación del término BIOMETRÍA, que apareció a fines del siglo XIX refiriéndolo al estudio de casos biológicos en que se usa instrumentos de la estadística.

II.2.-. Evolución de la econometría.

La evolución de lo que hoy se conoce como econometría tiene sus bases en la cada vez mayor disponibilidad de información estadística y la interfaz que se ha logrado hacer con la teoría económica. La aparición de la econometría como disciplina permitió que la teoría económica cambiara su metodología y, de ser un análisis de observación introspectiva, pasó al más puro empirismo histórico.

Algunos estudiosos como G.S. Maddala (1996:2) comentan que en opinión de Stigler en 1699 se identifica a Charles Devenant como pionero en el área, puesto que hizo y publicó ese año *el primer programa de demanda empírica*; otros consideran que su origen está en los trabajos realizados en dicho siglo XVII por Sir William Petty, quien escribió en 1676 y luego publicó en 1690 su obra la *aritmética política* o más recientemente, en 1758, con la *Tableau Économique* elaborada por el célebre médico francés Francois Quesney.

Alguna vez William Stanley Jevons aseveró que con la ayuda de la estadística podría llegar a conseguir desvelar las leyes de comportamiento de la economía de un modo tan preciso como en las ciencias exactas.[2]

En efecto, la econometría tiene sus bases en la estadística, sin embargo con la evolución de la estadística, matemáticas y de la propia teoría económica, ha abierto paso a la "econometría como un cuerpo de conocimiento independiente"[3]

La historia de la econometría es amplia y ha sufrido diversos cambios, que en la opinión de Julián Pérez (1998) pueden clasificarse de la siguiente manera:

- **Etapa I: Antecedentes**. Comprende desde las primeras expresiones económicas a través de las matemáticas, hasta 1914.
- **Etapa II: Desarrollos iniciales.** Abarca desde la publicación de H. Moore de Economic Cycles: Their Law and Causes, obra que muchos especialistas catalogancomo el primer trabajo econométrico formal, hasta 1930.
- **Etapa III: Formalización.** Se le llama así porque el 29 de diciembre de 1930 se fundó la Econometric Society en la ciudad norteamericana de Cleveland, a iniciativa de Charles Roos, Ragnar Frisc e Irving Fisher. Esta etapa llega hasta 1950.
- **Etapa IV. Extensión**. Va desde la publicación de la monografía número diez de la Cowles Comissión, en la cual se presentan las normas básicas de la investigación econométrica, etapa que se prolongó hasta 1970.
- **Etapa V. Diversidad de enfoques**. La cual se inicia con la publicación del Time Series Analysis de Box & Jenkins, que brinda un nuevo enfoque (modelos univariantes de series temporales, conocidos como modelos ARIMA) para explicar mejor (ya que los distintos modelos clásicos empezaron a fallar sistemáticamente en sus predicciones) el comportamiento y relación de las variables económicas en un sistema económico internacional agobiado por la crisis petrolera, el abandono del patrón oro y el cambio del flujo financiero que ocasionaron altos niveles de inflación y desempleo en economías completamente desestabilizadas.

La **Etapa I** se remonta hasta el siglo XVI con los trabajos aritméticos de Petty, King o Devenant, los cuales hicieron uso por primera vez de sistemas basados en hechos y cifras. Shumpeter, en su libro Historia del Análisis Económico (1994) describe a éste grupo como "una ilustración perfecta de lo que es la econometría,

[2]W. S. Jevons en Sánchez, González C., Métodos Econométricos, (1999)
[3]Sánchez (1999:4)

los econometrístas y lo que estamos tratando de hacer”[4] Este periodo se destaca por ser el de descubrimiento de las leyes en la economía, muy similares a las de las ciencias exactas.

Para el siglo XIX, con el nacimiento de la estadística moderna surgieron los trabajos de Galton, Edgeworth, Pearson, Cournot, Walras, Pareto, Jevons, entre otros. Se dice que Benito Pareto (1907), un estadístico italiano, fue el primero en hacer uso del método de regresión múltiple en la economía.

La **Etapa II** surge con Henry Moore en 1914, ya que fue éste el primero en estimar estadísticamente las relaciones económicas en el centro de análisis cuantitativo de la economía. A la par del trabajo de Moore y sus discípulos se comenzó a desarrollar el interés y estudio por los ciclos económicos basados en series de tiempo y la construcción de modelos econométricos para su estimación. Clemente Juglar fue el primero en estudiar los ciclos económicos, su mayor aportación es el Ciclo Juglar que es el ciclo de inversión con una duración de poco más de 11 años. Dicho tema fue de interés para otros economistas como Kitchin, Kuznets y Kondratieff, los cuales también aportaron sus ciclos.

El nacimiento de la **Etapa III** es paralelo al surgimiento de la Econometric Society y la aparición de la revista Econometrica. Esa sociedad en realidad es el conjunto de todos los esfuerzos realizados por figuras destacadas en la Teoría económica, estadística y matemáticas, de ahí la naturaleza multidisciplinaria de la econometría. El movimiento científico alrededor de la econometría tuvo su momento cumbre con la creación de la Sociedad Econométrica Internacional el 29 de diciembre de 1930 en Cleveland, Ohio bajo la batuta de Fisher, Frisch y Roos.

El primer número de la revista Econometrica se publicó en 1933, ésta revista constituyó un medio de comunicación entre los econometrístas del mundo.

Se estableció, además la Cowles Commision, la cual se erige como el centro de la ortodoxia econométrica. Apareció también en Inglaterra el Department of Applied Economics en Cambridge en 1939 bajo la dirección de Keynes, sin embargo comenzó a operar hasta finales de la segunda guerra mundial.

Esta etapa es fundamental en la dotación de un conocimiento generalmente aceptado.

Dentro de toda la euforia por el nuevo acontecimiento, la preocupación de Frisch seguía siendo no solo la crítica de Keynes, sino los problemas de multicolinialidad y errores de medición y para estudiarlo inventó el método de mapas de racimo, el cual no aporta mucho a la profesión. Bajo la misma preocupación se manejó Koopsman el cual hizo mayor hincapié a la utilización de modelos estocásticos.

[4] Shumpeter, en http://ftp.iza.org/dp2458.pdf

Para 1959 inicia la **Etapa IV** o de extensión con la publicación de la monografía número 10 de Cowles Commision bajo el título de Stadistical Inference in Dynamic Economic Models, en la cual se establecen las normas básicas de la elaboración de las investigaciones econométricas, dicho esfuerzo se vería cristalizado hasta la monografía número 14, publicada en 1962, con el título de Studies in Econometric Methods.

Es dentro de ésta etapa, a principios de los años cincuentas, que se comienza a considerar a la econometría como un cuerpo teórico maduro dentro del campo formal, es por ello que comenzaron a aparecer los primeros libros y textos de econometría, como A text book of Econometrics, publicado por Klein, Econometric Methods de Jonhston, el cual es ya un clásico que ha servido de guía para la realización de manuales posteriores, ya que, entre otras cosas, analiza los problemas de identificación, estimación y propiedades para muestras pequeñas en los modelos multiecuacionales.

Es en esta etapa cuando los españoles comienzan a escribir textos en los que se destaca la exposición ordenada de los avances más relevantes de la econometría. La primera obra publicada fue en 1962, bajo el nombre de Fundamentos y Posibilidades de la Econometría del profesor A. García Barbacho. Para 1963 A. Alcaide publicó Conferencias sobre Econometría y Métodos Estadísticos, en el cual, además de mostrarse los modelos econométricos más actuales, incorporó la elaboración de tablas de input-output y números índices, entre otros.

Entre los aportes teóricos de ésta etapa sobre sale el método de los mínimos cuadrados trietápicos (MCT) expuesto por Theil y Zellner en la revista Econometrica en 1962, dentro del artículo “Three-Stage Last squares. Simultaneous Estimations of Simultaneous Equations”.

En resumen, en la Etapa IV, se perfeccionan los métodos de estimación disponibles hasta ese momento, se desarrollaron alternativas para sustentar los problemas derivados del incumplimiento de las hipótesis a priori. Es en ésta etapa de extensión se llevan a cabo los primero esfuerzos para modelar, utilizando sistemas de aplicaciones empíricas, así como la publicación de múltiples obras que vienen a reforzar la etapa de madurez que tomaba la econometría.

La **Etapa V**, es la etapa de diversidad de enfoques, la cual comienza en 1970, con la temática conocida en los manuales como “Teoría de la Econometría”, así como la generalización del uso de la prueba Durbin Watson para la detección de autocorrelación en aplicaciones empíricas y finalmente los estudios sobre modelos con retardos distribuidos, pero sobre todo, sobre modelos dinámicos. Sobre el tema de los retardos, los libros realizados sólo fueron una continuación de los estudios llevados a cabo por Koyck en la década de los cincuentas.

El cambio suscitado en el campo de la econometría hacía mediados y finales de la década de los setentas, está relacionado con el incremento en el precio de los energéticos y la crisis que le siguió y en la que se vieron involucrados la mayoría de los países, en los cuales se incrementó la inflación y el desempleo. Para esos fenómenos, la teoría dominante no fue capaz de proporcionar una explicación. Para la econometría lo anterior representó un problema, ya que todos los modelos estructurados hasta el momento comenzaron a presentar fallas sistemáticas en sus predicciones, por ello se puso en duda la capacidad predictiva, baluarte en todos los modelos hasta esa fecha.

Antes de entrar a la década de los ochenta, R. Lucas expuso en su trabajo Econometric Policy Analysis: A Critique, la inadecuación de los modelos econométricos para adaptase a políticas alternativas, ya que al modificarse las políticas, se modifica por ende la estructura de los modelos, invalidando el proceso de contrastación. Esta crítica fue de las más importantes en cuanto a la incapacidad e ineficiencia de los modelos de la época.

En la década de los ochenta se intentó superar la ineficiencia de los modelos estructurales a nivel mundial, en lo referente a su aspecto predictivo. Las críticas esbozadas durante la década de los setenta tomaron importancia durante la época de los ochenta. Entre las críticas de mayor impacto teórico se pueden mencionar la de David Hendry, Edward Leamer y Chistopher Sims.

Hendry tomó como estrategia para la modelización la especificación rigurosa del modelo antes de estimarlo y validarlo. La diferencia principal entre ésta forma de modelar y la tradicional es el énfasis puesto en la especificación del modelo. De tal forma que, dentro del nuevo modelado, tras un diagnóstico desfavorable, lo recomendable es poner en marcha una nueva especificación del modelo. Cuando la respecificación del modelo ocurre, suele incluirse nuevas variables o se reformulan las anteriores.

Leamer, por su parte, publicó en 1983 “Let´s Take the Con out of Econometrics”, en el que pone de relieve la subjetividad de la modelización econométrica y sus resultados, recomendó entonces, que dicha subjetividad e incertidumbre quedara recogida en un análisis específico al que denominó “análisis de valores extremos”.

Sims en 1980 publicó en la revista Econometrica un artículo llamado Macroeconomic and Reality, en el que el tópico inicial fue su escepticismo sobre las hipótesis establecidas para los modelos econométricos, así como su alternativa en la que propone que sean los datos los que lleven a la especificación concreta del modelo. La tesis que trabajó Sims se basó en que la teoría económica no proporciona la especificación de ningún modelo, así como las variables endógenas o exógenas de cada ecuación, a lo que llamó hipótesis de exogeneidad y restricciones de identificabilidad, respectivamente. En resumen, ésta propuesta es metodológica, sin contenido teórico, en el que todas las variables dependen de todas, según la estructura de los retardos que se puede especificar a través de datos.

Aunado al análisis que hizo con respecto al modelado econométrico, Sims abordó otra problemática, la de los instrumentos de política económica y las decisiones que se toman con base a éstos. La decisión de política económica se puede dar desde dos enfoques de evaluación de ésta. El primero es el enfoque de instrumentos-metas, llamado prospectivo, el cual consiste en proponer acciones alternativas que le competan al sector público, para analizar los efectos sobre variables endógenas, de tal manera que, bajo la norma de eficiencia, sea más simple elegir la opción adecuada. El segundo enfoque es el de metas-instrumentos, nombrado también introspectivo, el cual se basa en la coherencia de las metas de política económica planteadas para evaluar las estrategias a realizar. Cuando la estrategia fue seleccionada, entonces se somete a evaluación en términos de factibilidad, sin embargo, los gobiernos no encuentran esa factibilidad, lo cual representa un problema dentro de éste enfoque.

Es a partir de ese cambio estructural y la incapacidad de los modelos de explicar la nueva realidad económica que surge el desarrollo que marcó todos los estudios posteriores: el sesgo funcional. Este hecho marca un parte aguas en la econometría ya que ahora su objetivo se centra en la obtención de aproximaciones empíricas buenas en dentro del conjunto explicación-predicción-escenarios, de tal modo que ahora las aplicaciones prácticas se comienzan a juzgar por su capacidad de simulación, predicción y explicación.

Bajo ese contexto, surgen críticas a la modelización tradicional representada por los modelos macroeconométricos de ecuaciones simultáneas que sirvieron como estímulo para la creación de nuevas especificaciones y técnicas para abordar el análisis económico aplicado, tal y como el que se presentó en los primeros trabajos de Box y Jenkins sobre modelos univariantes de series temporales (modelos ARIMA). En ésta etapa, además de las aportaciones teóricas, la incorporación de ordenadores electrónicos hizo posible la aplicación de las nuevas tecnologías sobre modelos económicos.

La obra de Box y Jenkins aunada a una mayor disponibilidad de información estadística y la facilidad de cálculo otorgada por el desarrollo de la microelectrónica propició una nueva metodología para el tratamiento de series de tiempo: los modelos vectoriales autorregresivos (modelos VAR), análisis de cointegración, modelos estructurales de series temporales, entre otros. Retomaron importancia los modelos de alisado, los cuales se habían dejado de utilizar durante la década de los cincuentas y cuyo fundamento recae en la idea de que hay posibilidad de explicar el comportamiento futuro de una variable en el corto plazo, simplemente conociendo su comportamiento pasado. Otro análisis que tomó fuerza fue el espectral o frecuencial de procesos estocásticos, el cual se basa en la descomposición de las series históricas en componentes de frecuencia alta o rápida y de frecuencia baja o lenta.

Los modelos ARIMA, representan, desde la óptica de la predicción, una generalización de los modelos de alisado, aunque con un avance notable sobre estos. La capacidad predictiva a corto plazo de los modelos ARIMA se mostraría superior a los macromodelos, los cuales comienzan a tomar mayor fuerza a partir de la década de los setenta, ya que son éstos los que servirían de base para tomar decisiones de política económica.

El argumento respecto a que los modelos ARIMA presentaban superioridad sobre los modelos estructurales o macromodelos fue que los primeros, realizados bajo la metodología de Box-Jenkins, reaccionaban inmediatamente, a los cambios importantes, es decir, tenían mayor flexibilidad de adaptación que los modelos estructurales, los cuales tardaban en acostumbrarse a cualquier marco económico. Fue hasta 1974 cuando las discrepancias entre modelos quedaron atrás con la aparición de la obra de A. Zellner y F. Palm, Time Series Analysis and Simultaneous Equation Econometric Models, en la que se plantea la integración de ambos enfoques, es decir, el análisis causal con el análisis de series de tiempo. Una forma de integrarlos se dio en 1981 con la obra Modelling Multiple Time Series with Applications bajo el modelo VARMA.

II.2.1.-México

En nuestro país los modelos estructurales comenzaron a utilizarse a principios de la década de los setentas bajo la dirección de Klein y Adams como proyectos de tesis doctoral, un ejemplo de esto es la tesis de Abel Beltrán.

Los modelos estructurales sirvieron como punto de partida para la evaluación de las políticas económicas aplicadas por los gobiernos mexicanos de los años setentas y ochentas, lo cual llevó a que se tuviera una estrecha relación entre académicos y gobierno, posteriormente el sector privado encontró utilidad en la información obtenida de los modelos macroeconómicos, por lo que se volvieron cada vez más prolíferos, de tal forma que los modelos mexicanos han sido elaborados por el sector privado, público y universidades.

El primer modelo académico mexicano corrió a cargo de la Universidad Nacional Autónoma de México (UNAM), coordinado por el maestro David Ibarra en 1970 y cuyo objetivo es el análisis de las propuestas de la política económica para dinamizar el mercado laboral durante el sexenio de Luis Echeverría. El modelo es estimado con mínimos cuadrados ordinarios (MCO), consta de 19 ecuaciones y su enfoque es estructuralista, el periodo de análisis es de 1950 a 1966 en series anualizadas. El mérito de este trabajo consiste en que es el primero en tratar problemas económicos estructurales, así como la dependencia cada vez más determinante del exterior, lo cual además es determinante después de 40 años.

Posteriormente, en 1976, surgió el modelo de Clavijo, el cual tiene por objetivo la estimación y análisis coherente de la economía mexicana de ese entonces, así como un pronóstico para 1976 a 1977. El modelo se estima por MCO y rezagos polinomiales escalonados (ALMON), el periodo comprendido en el modelo es entre 1965 y 1975, de forma trimestral. Consta de 9 ecuaciones estocásticas y 22 identidades, el enfoque es monetarista en lo referente al gasto de gobierno y sus afectaciones a la demanda agregada.

En 1979 se publicó el modelo de planeación hacendaria, el cual constituye el primer modelo realizado por una dependencia del Gobierno Federal, el cual, como los anteriores, tiene como objetivo el análisis y evaluación de las políticas económicas, con el fin de realizar un pronóstico para la toma de decisiones por parte de los funcionarios públicos. El modelo es estimado entre 1959 y 1977 de forma anual por MCO. Se conforma de 38 ecuaciones de comportamiento bajo el enfoque keynesiano, vigente aun en la época, del modelo IS-LM. Se buscaba resaltar la política del gasto y financiamiento público, tomando en cuenta el precio al alza del petróleo y las exportaciones de éste.

La década de los ochenta también fue prolífera para los modelos mexicanos, que además fueron fruto de la ola de cambios que se vinieron dando en el campo de la econometría, como es el caso del modelo de Ruffat surgido en 1981 y cuyo fin es la incorporación de una metodología para establecer una función de bienestar, como meta, a partir de la cual se determinarían los instrumentos de política económica a seguir, tal y como lo planteó Sims en su enfoque introspectivo.

En 1984 surgió el segundo modelo académico mexicano, el cual fue desarrollado por el Centro de Investigación y Docencia Económica (CIDE) y que lleva por nombre: modelo Modem, su objetivo es tipificar la dinámica económica mexicana, estimando el periodo anual de 1960 a 1982 bajo el enfoque keynesiano-estructuralista.

El modelo Aspe-Jarque se dio a conocer en 1985 con una estimación trimestral comprendida entre 1972 y 1982 y cuyo objetivo es probar el cumplimiento de las expectativas racionales para la inflación y el Producto Interno Bruto (PIB) por medio de la razón de verisimilitud.

En ese mismo año se dio a conocer el primer modelo que implementó el cambio estructural por medio de la prueba Chow. El periodo de estimación comprende de 1950 a 1980 y lleva el nombre de modelo Amieva Huerta.

El modelo Ricardo Lago conocido a partir de 1991 fue el primero que destaca la importancia de la programación financiera, el cual es estimado para el periodo anualizado de 1960 a 1983. La contribución de este modelo consiste el perfeccionamiento de la metodología tradicional, al permitir la determinación de la balanza de pagos, actividad económica, meta inflacionaria y déficit público al mismo tiempo.

En la actualidad se han hecho modelos para la toma de decisiones dentro de las instituciones de gobierno como es el caso del Instituto Nacional de Ecología que desde 2007 cuenta con el modelo de renta de la tierra, el cual se basa en la premisa de que cualquier tierra puede tener diversos usos que compiten entre sí, dejando una renta al dueño de la tierra. El modelo se compone de siete variables interactivas entre las que destaca la población, ejidatarios y la ganadería; y se ha venido estimando desde 1996.[5]

La Cámara de Diputados realizó un estudio que serviría para la toma de decisiones sobre la sostenibilidad de las finanzas públicas durante el periodo comprendido entre 1997 y 2007, en análisis trimestral. El modelo está basado en la premisa de que la sostenibilidad de la deuda pública se da mediante la Restricción Presupuestaria Intertemporal del Gobierno, esta idea está estrechamente ligada a la idea de que el déficit gubernamental compromete a las generaciones futuras. El modelo está formado por dos ecuaciones, las cuales constan de cuatro variables cada una.[6]

Es la misma Cámara la que en 2009 realizó un estudio basado en un modelo macroeconométrico que sirvió para tomar mejores decisiones a cerca de los pronósticos a largo plazo del IVA, ya que fue en este año cuando se tomó la decisión de aumentar éste impuesto un 1%. La serie de tiempo a analizar es del primer trimestre de 1990 al segundo trimestre de 2009, tomando en cuenta el cambio de año base a 2008. La estimación se realizó en tasas de crecimiento, teniendo cinco variables explicativas. La finalidad de este modelo es mostrar que en el largo plazo un aumento en el producto, como se esperaba para 2010, traería consigo un aumento en la recaudación y sería sostenible. Cabe destacar que éste estudio se tomó como base para tomar la decisión de aumentar la recaudación por concepto de IVA. [7]

El Banco de México ha realizado, por su parte un sinnúmero de estimaciones a través de modelos macroeconométricos, uno de los que más salta a la luz es el publicado en 2004 que trata de pronosticar la formación bruta de capital privado en nuestro país, para ello se valió de un modelo autorregresivo de rezagos distribuidos (ADL) en términos de modelos de corrección de error. Cuenta con diez y seis variables las cuales están tasadas, es decir, son logaritmos. La serie de tiempo utilizada para éste trabajo abarca del primer trimestre de 1980 al tercer trimestre de 2002.[8]

---

[5] http://www2.ine.gob.mx/publicaciones/gacetas/464/procede.html

[6] www.cefp.gob.mx

[7] http://www.cefp.gob.mx/intr/edocumentos/pdf/cefp/2009/cefp0972009.pdf

[8] http://www.banxico.org.mx/documents/%7B1EDA8A25-618D-690F-A73B-DD4F93AB8D9C%7D.pdf

II.2.2.- Los modelos macroeconómicos más conocidos.

II.2.2.1. Norteamericanos.

Se dice que en 1998 apareció publicado al mismo tiempo que se inauguró un sitio en Internet gratuito con la versión electrónica, el MME para la economía de EEUU con el nombre de US Model, de Ray Fair, profesor de la Universidad de Yale. Entre otras aplicaciones, es útil como simulador al contener información para crear escenarios en sectores claves como: los hogares, las empresas, el sector financiero, el sector gobierno federal, el sector gobierno local, y el sector externo.

II.2.2.2.-. Los españoles.

Destacan el modelo Wharton –Universidad Autónoma de Madrid, (WUAM), que existe desde 1981. También, el Modelo de investigación y simulación de la economía española (MOISSES) y, el que se publicó en 1994, intitulado: "Modelo de Aproximación Trimestral, MAT.

Indudablemente que deben de existir otras experiencias interesantes en el terreno profesional y en el académico sobre el uso de esta disciplina en el análisis económico; sin embargo, yo me concretaré a estos con los cuales espero haber despertado la inquietud de los lectores para que hagan su propia investigación y enriquezca este espectro tan valioso de esta disciplina.

II.- 3.- Definición de econometría.

II.3.1.-Original y su evolución.

Etimológicamente significa *medición económica*; posteriormente se le ha relacionado con el *análisis de regresión y de correlación*, el cual amplió mucho su horizonte de aplicación porque el investigador no sólo estuvo en condiciones de medir a las variables, sino también de relacionarlas entre sí, es decir, *de formular teorías económica*s, motivo por el cual se leconoce como la disciplina que proporciona la metodología para expresar estadísticamente las relaciones que existen entre un fenómeno bajo estudio y las variables que lo explican.

II.3.2.- Aportación de Galdon y Pearson.

El concepto de análisis de regresión fue introducido por Francis Galton quien estudió las estaturas de un número de padres de familia y de sus hijos; observó *cierta tendencia:* los padres de estatura alta tenían hijos altos y los padres de estatura baja tenían hijos bajos (Gujararti, 2003: 17) y que la *estatura promedio* de los hijos nacidos de padres de una cierta estatura tendía a moverse o *regresar* hacia la estatura promedio de la *población total*. Su hallazgo reveló que "la estatura de los hijos inusualmente altos o de padres inusualmente bajos tiende a *moverse* hacia la estatura promedio de la población total.

Sus hallazgos también conocidos como *Ley de Regresión Universal* fueron corroborados por su amigo Kart Pearson con una muestra de mil individuos cuyas estaturas registró, analizó y encontró "que la *estatura promedio* de los hijos de un grupo de padres alta era menor que la estatura de sus padres y la *estatura promedio* de los hijos de un grupo de padres de estatura baja era mayor que la estatura de sus padres, generando un fenómeno mediante el cual los hijos altos e hijos bajos, "regresaban" por igual hacia la *estatura promedio* de todos los hombres". En palabras de Galton, se trataba de una "regresión hacia la mediocridad".

Dicha variación en la estatura promedio de los hijos dada la estatura de los padres, es una *estimación* que se obtiene con la recta de regresión, cuyos *valores de los parámetros* de la ordenada al origen y de la pendiente de la variable independiente (la estatura de los padres) se obtienen con el método de ***Mínimos cuadrados.***Así, si le damos valores a la variable independiente (al menos dos) se puede graficar y a esta línea se le conoce como ***Recta de Regresión***, misma que expresa "la forma en que *el promedio* de la estatura de los hijos aumenta conforme lo hace la de los padres" (Gujarati, 2003: 19). Dicho en otras palabras, la recta de regresión indica en cada uno de sus puntos que *a un valor dado* de la estatura de los padres le corresponde una gama de estaturas de sus hijos *pero que dicha gama o distribución de estaturas tiene un promedio*, que es el valor representativo de todos ellos (valores altos y bajos de las estaturas de los hijos). Al siguiente valor de la estatura de los padres también le corresponde un rango de estaturas de los hijos; ese rango o distribución de estaturas también tiene un promedio, que es el siguiente punto de la recta de regresión. Así podemos ir obteniendo sucesivos valores promedio de la recta, los cuales son *estimaciones de la estatura* de los hijos (variable dependiente) obtenidas por el método de mínimos cuadrados. En resumen, los puntos de una recta de regresión son valores promedio de la estatura de los hijos dada la estatura de los padres. Todo este marco teórico de estimación se verá en detalle enseguida.

En este sentido ahora es conveniente indicar que un fenómeno económico se estudia a través de la observación. Su comportamiento se registra y evalúa preferentemente con datos cuantitativos (en ocasiones con cualitativos, expresados a través de variables llamadas categóricas, dicotómicas, ficticios o dummy). La matemática permite expresar su comportamiento a través de ecuaciones que pretenden describir determinada teoría económica, misma que en turno permite a la estadística indicar si es o no verídica. Así, una vez que se expresa la teoría económica en forma *uniecuacional o multiecuacional*, la estadística proporciona los métodos para corroborar sí se prueba o no con rigor técnico.

II.3.3.- Definición apropiada para captar fácilmente la importancia de la econometría.

Mi opinión es que “es una disciplina con la que una *teoría económica* se puede describir o expresar *matemáticament*e y probarse con los métodos de la estadística inferencial”. En esta perspectiva se puede decir que los econometristas tienen como meta usarla para explicar e interpretar cuantitativamente una teoría sobre un fenómeno bajo estudio, cuya realidad o universo estadístico conoce parcialmente a través de una muestra que utiliza para describirla en términos estadísticos, es decir, para inferir a través de los valores muestrales los valores o características estadísticas de los datos que integran la población o universo estadístico.

El experto Julían Pérez (1985: 10) comenta que esta conceptualización de la econometría se registró en 1959, año en que también se enfatizó su función de cuantificar “el impacto de una variable sobre otra, así como para predecir acontecimientos futuros o aconsejar qué política económica debería seguirse cuando se desea alcanzar determinado resultado (Valvanis, 1959)”.

Así pues, considero que esta definición ha pasado la prueba del tiempo, se ha verificado que sigue siendo la expresión sencilla sin alambicamientos conceptuales de esta disciplina; en particular creo que es fácilmente entendido su significado y alcance entre los personas que inician su aprendizaje de la misma, dado que de inmediato se visualiza que se moverá en un escenario en que los actores son la teoría económica, las matemáticas y la estadística y, cuya instrumentación ahora cuenta con el valioso apoyo de las NTIC.

II.3.4.- Objetivo actual.

De ahí que la econometría tenga como objetivo la determinación empírica de las leyes económicas. En ese contexto es que también se le utiliza para probar teorías o hipótesis económicas, cuya verificación econométrica las convierte en acervo científico.

Como puede observarse el análisis de regresión amplió considerablemente su ámbito de trabajo ya que pasó de meras mediciones o cuantificaciones de un fenómeno económico (variable dependiente) a su relación o vinculación con otro(s) fenómenos (variables independientes), permitiendo con ello formular relaciones, hipótesis o teorías como mejor convenga a los objetivos del investigador.

II.-4.-Alcance y limitaciones de la econometría.

II.4.1.-Alcance.

1.- Es de gran ayuda para modelar políticas micro y macroeconómicas, para hacer análisis de estructura, de predicción y para evaluar dichas políticas, en virtud de que es la integración de una teoría económica con las matemáticas y las técnicas estadísticas, que fundamentan sólidamente a la primera.

2.- Es muy útil porque con su metodología se puede probar la relación hipotética que se establezca entre las variables de interés para el investigador. Cuando la *relación estimada* no pasa las pruebas estadísticas, es necesario modificar la relación hipotética (teoría económica) entre las variables bajo estudio. Si pasa dichas pruebas, se cuenta con un sustento metodológico para decir que la relación existe entre las variables.

3.-Sabiendo que *existe la econometría teórica y la aplicada*, en el análisis de regresión la econometría teórica se complementa adecuadamente con la aplicada o empírica, ya que la primera proporciona los métodos para *medir*la relación económica que tiene la variable dependiente con la independiente o explicativa. La segunda, cuantifica dicha relación con el valor del coeficiente de la variable independiente, así como los hallazgos y problemas que surgen cuando se estudian relaciones en campos específicos de la economía, como la teoría de la inversión, del empleo, del consumidor, etc.

4.- En el análisis de regresión dichos coeficientes determinan el efecto que tiene la variable independiente sobre la dependiente, lo cual permite hacer análisis de estructura porque expresa la conexión que existe entre las variables en estudio y, a su vez, pronosticar el impacto que tendrá en el horizonte de planeación que se establezca entre ellas.

5.- Proporciona la metodología para verificar empíricamente una teoría o hipótesis, es decir, es el medio para admitir dicha teoría dentro de la investigación científica y para la toma de decisiones fundamentada en materia de políticas económicas.

II.4.1.1.-Ejemplos:

II.4.1.1.1.- Aplicaciones en políticas macroeconómicas.

Supóngase que para dinamizar el crecimiento de la economía mexicana el Ejecutivo Federal a través de la Iniciativa de Ley de Ingresos propone al Congreso de la Unión una reducción en los principales impuestos federarles: IVA e ISR para incentivar el PIB. La pregunta que surge de inmediato *¿en cuánto aumentará la inversión con este incentivo fiscal y por ende, en el PIB y el Consumo?*

Para aprobar dicha medida fiscal de estímulo y conocer el impacto en la inversión y, en turno, el efecto de ésta en el crecimiento del PIB y del Consumo, en el Congreso de la Unión se analiza con el análisis de regresión su impacto sobre el crecimiento del PIB y, el Consumo

Para ello se selecciona una muestra constituida por datos de series de tiempo del PIB y del Consumo. Se establece la hipótesis de que el Consumo depende del

PIB y, aplicando el *método de mínimos cuadrados* para encontrar la ecuación de regresión que muestre el coeficiente o pendiente de la variable independiente (PIB), cuyo valor *expresa la propensión marginal a consumir (PMC)* y que de acuerdo con la teoría macroeconómica, además de indicar cuánto se dedicará al consumo por cada unidad de aumento en el PIB, sirve para determinar el *multiplicador de la inversión (M),* así:

M=1/1-PMC

Así, supóngase que *PMC* es igual a 0.72, tiene dos interpretaciones:

1.- Indica que por cada aumento de *una unidad* (un peso) en el PIB el consumo nacional aumentará en 72 centavos; y

2.- Que M será igual a 3.57, indicando que un aumento (o reducción) de un peso en la inversión inducirá un incremento (o reducción) de 3.57 veces en el PIB.

Cabe señalar que en este caso el cálculo clave es el de la PMC en virtud de que M se obtiene con su valor. Como puede observarse la PMC proporciona valiosos elementos para la adopción de una política pública, en este caso, una política fiscal que incidirá en el crecimiento del PIB, del Consumo y finalmente, en el empleo y calidad de vida de la sociedad.

Asimismo, el crecimiento futuro de estas variables macroeconómicas (PIB y Consumo) se puede estimar aplicando a la *ecuación de regresión el método de predicción,* como se verá más adelante.

II.4.1.1.2.- Aplicaciones en la política de la empresa.

Una vez que sabemos que mediante el análisis de regresión podemos relacionar dos o más variables, donde una de ellas es la dependiente y el resto son las independientes o explicativas del comportamiento de la primera (dependiente), así como pusimos el ejemplo a nivel macro, ahora podemos aplicar el análisis de regresión a nivel empresarial. Para ello primero debemos *definir la política de la empresa que se desea implementar, digamos la de ventas*. Con esas referencias enseguida se debe definir qué variable será la dependiente (ventas) y las que suponemos que las expliquen, quizás: publicidad, mercadotecnia, etc.
Así, el siguiente paso es buscar sus datos y usarlos para calcular la ecuación de regresión, cuyos parámetros (de la publicidad y de la mercadotecnia) indicarán la relación que tienen estas variables independientes con la dependiente (ventas). Con esos resultados, podemos construir un escenario de planeación, digamos a cinco años. Este pronóstico de ventas en función de la publicidad y la mercadotecnia, nos permitirá formular la política de ventas de la empresa en ese lapso de cinco años, en la que se describan las acciones a realizar para que sea competitividad de alta productividad y rentabilidad.

II.4.2.- Limitaciones

1.- Al expresarse la relación entre las variables mediante un modelo econométrico, y, por consiguiente, al considerarse éste último como una representación sencilla de la realidad, siempre quedará la duda sobre si la estimación fue o no correcta, situación que obliga a revisar cuidadosamente su especificación.
2.- En este contexto, si en el análisis de regresión se usa el *método de mínimos cuadrados* para *estimar* el valor de los parámetros de las variables independientes, si no se cumplen los supuestos teóricos que sustentan dicho método, las *estimaciones* pueden no tener las propiedades que sustentan la solidez de las pruebas de hipótesis y por consiguiente, inducir a la toma de decisiones equivocada: por ejemplo, aceptar o rechazar la hipótesis o teoría económica.
3.- Al usarse una muestra de datos para expresar la relación entre las variables, siempre quedará la duda sobre el grado de representatividad que tenga la muestra de los datos, aun cuando se especifique la *probabilidad*de que el error de muestreo sea diferente del error permitido.
4.- Al basarse la econometría en el análisis de regresión suele decirse que dicho análisis identifica la *relación causal* entre las variables independiente y dependiente; lo anterior no siempre es posible porque los métodos utilizados no permiten hallar una causalidad absoluta, entre otras razones porque,“ las relaciones entre las variables económicas son inexactas y algo erráticas”(Salvatore, 1999:4); yo agregaría que si dicha relación de causa-efecto no se especificó en la teoría económica, entonces está relación es un mero ejercicio estadístico o como dice Surignac (2001) “ es una relación que aflora por casualidad y no por causalidad.

Por la importancia del tema agreguemos lo siguiente: Comenta Gujarati (op.cit, 22) que aún cuando el análisis de regresión como lo visualizaron Galdon y Pearson: *relación de dependencia de una variable Y de otra X*, en opinión de varios estudiosos eso no siempre indica una relación de causa-efecto; para ello cita a Kendall quien indicó que “ Una relación estadística, sin importar que tan fuerte y sugestiva sea, nunca podrá establecer una relación causal. Nuestras ideas de causalidad deben venir de estadísticas externas y, en último término, de una u otra teoría”.

Para ilustrar esta relación usa el ejemplo siguiente: “El hecho de que se trata al producto de la cosecha como dependiente de la lluvia, ello se debe a consideraciones no estadísticas”, yo intuyo que quiere decir que ello se debe atribuir a la teoría que previamente se ha establecido, cuya modelización matemática hace depender el producto de la cosecha de la lluvia, independientemente de que al correr la regresión entre estas dos variables los resultados indiquen que existe una fuerte relación estadística, en otras palabras: que no hay una relación de causa-efecto sino una relación establecida apriorísticamente: la teoría económica modelada matemáticamente.

II.4.3.- Limitaciones de la metodología.

Julián Pérez (1998) comenta que dentro de las imperfecciones intrínsecas de sus técnicas, suelen mencionarse algunas como las siguientes:

a).-Las técnicas iterativas aplicadas “no siempre garantizan” que se obtenga una solución óptima;
b).- Las pequeñas perturbaciones que se presentan al inicio pueden incidir significativamente en los resultados finales;
c).- La inferencia estadística aplicada durante el “proceso de estimación tiene un sesgo de origen”, en virtud de que una vez obtenido el modelo se realizan dostipos de evaluaciones, que aún cuando están estrechamente vinculadas, se les maneja “como fenómenos independientes”. Ellas son el análisis de significación de sus parámetros poblacionales y las características estadísticas que presumiblemente tiene la distribución de la perturbación aleatoria;
d).- Comenta que continúan utilizándose técnicas que a pesar de que se acepta que tienen imperfecciones teóricas, no se abandona su uso. Ellas son: “los niveles de significancia de una distribución *t-student* para contrastar coeficientes de variables con tendencia estocástica, estos se continúan usando”, cuando se ha demostrado que dichos niveles de significativilidad no son válidos”, para ello se apoya en MacKinnon (1987);
e).- En esta tesitura señala que están los métodos de estimación como el MCO: Mínimos Cuadrados Ordinarios; que se usa en modelos simultáneos aún cuando desde hace tiempo se conoce que son mejores los “métodos de información completa o de los estimadores de máxima verosimilitud”.

Al respecto, conforta su opinión en el sentido de que “las técnicas econométricas son aproximaciones científicas en proceso continuo de perfeccionamiento: posiblemente en un grado de maduración previo al de otras ciencias duras, pero no por eso en una escala inferior del rigor científico, sino por el contrario, situado en una medida paralela del contexto de las ciencias praxeológicas en que se enmarca”.

6.- En este contexto es que Wooldridge (2009:12) comenta <u>en el caso de las series temporales</u> que “ en la actualidad es muy aceptado que los ejemplos tradicionalmente empleados para ilustrar la aplicación de los métodos econométricos a las series de tiempo son inadecuados o erróneos”, y que para superar esta limitante es necesario estudiar a fondo en una sección especial los temas relativos a tendencia, persistencia, dinámica y estabilidad, ya que el estudio de las series temporales es muy complejo porque se observan en sus datos *trayectorias de carácter persistente* en las series económicas que es necesario tomar en cuenta para mejorar la estimación de sus parámetros.

7.- Derivado de la complejidad de la realidad económica, algunos estudiosos de la *econometría señalan que ésta es un arte y una ciencia*, dado que para que brinde resultados confiables es necesario saber especificar la teoría económica que fundamenta el modelo, los métodos a utilizar y los datos para obtener y estimar correctamente las relaciones entre las variables en estudio; situación que en mucho depende de la experiencia del investigador, de su formación académica, de su intuición, de su buen juicio y familiaridad con el fenómeno de interés para él, entre otros factores.

II.5.-Perspectivas

En el siglo XXI, el instrumental teórico sobre análisis econométrico ha sido fortalecido por un desarrollo exponencial de las nuevas tecnologías informáticas y de las comunicaciones, NTIC, las cuales han aportado poderosas herramientas computacionales, bases de datos, técnicas y conceptos estadísticos que han permitido el uso de sistemas de resolución muy complejos en periodos muy cortos. Asimismo, los lenguajes de programación "amigables" han facilitado a los investigadores la posibilidad de generar sus propias herramientas de cómputo, lo que ha diversificado enormemente las posibilidades de aplicación teórica a la econometría empírica.

En este contexto es interesante indicar que dentro de la investigación , ahora como antes *se continúan analizando profusamente las propiedades estadísticas de los estimadores, ya que ellas son las que dan la calidad del ajuste estadístico de los valores de los estimadores a los valores de los parámetros poblacionales*, tanto con la econometría tradicional como con la de series de tiempo de fenómenos económicos, mediante la aplicación de las metodologías de los temas de cointegración y de raíces unitarias, cuyo origen se localiza en los modelos ARIMA (Box & Jenkins) y en el concepto de regresión espuria (Granger & Newbold, 1974), así como también la relación de equilibrio que debe tener la estructura de una serie de tiempo entre el corto y el largo plazo (con el mecanismo de corrección de errores) para hacer mejores predicciones.

Estos son algunos de los conocimientos frontera que han modificado el estado del arte econométrico.

III.ASPECTOS FUNDAMENTALES DE LA ECONOMETRÍA.

III.1.-Marco teórico.

Si por marco teórico se entiende el conjunto de ideas o conceptos que se argumentan en pro de un objetivo determinado, entonces en econometría es fundamental definir, describir e ilustrar el cuerpo teórico que es fundamental conocer para estar en condiciones de saber analizar e interpretar los resultados que se obtienen con la aplicación de su metodología a cierta teoría económica de interés del investigador.

Con el fin de contar con los conocimientos básicos requeridos para poder hacer análisis e interpretación apropiada y suficiente de los resultados que se vayan obteniendo con los cálculos de los valores de los estimadores, de la bondad de ajuste de los mismos a los valores de los parámetros poblacionales, del grado de asociación o relación que tengan entre si las variables explicada con la explicativa, de las pruebas de hipótesis que se realicen para verificar si se constata la teoría económica, etc. he juzgado conveniente establecer un marco conceptual .que sirva de referencia para interpretar correctamente los resultados que se obtengan y para evaluar si es suficiente aplicar uno u otro método de trabajo. Así, dado que el método de MCO: Mínimos Cuadrados Ordinarios continúa siendo uno de los métodos econométricos preferidos para obtener las *estimaciones* de las pendientes de los coeficientes de las variables independientes introducidas en el modelo lineal de regresión, es importante comprobar que no se violaron los supuestos establecidos para su uso, es decir, que se satisfacen, de manera que es necesario introducir la teoría que subyace o justifica su aplicación, etc.

Por lo anterior es que se ha juzgado necesario introducir los conceptos teóricos que sustentan los planteamientos de la relación entre variables aleatorias, así como para que sirvan para responder conceptualmente a muchos problemas que suelen presentarse cuando se hace econometría empírica.

Lo anterior no significa que al momento de desarrollar los temas en sus capítulos correspondientes ya no se vayan a utilizar los conceptos teóricos que fundamentan la aplicación de la metodología apropiada, no, eso no sucederá, por supuesto que se meterán ya que constituyen el eje rector de los desarrollos e interpretaciones que allí se hagan. La diferencia con el marco teórico de este capítulo es que éste constituye el basamento conceptual fundamental, por lo que allá será ampliado y enriquecido cuando proceda con las ilustraciones, derivaciones y análisis casuísticos que se realicen sobre resultados específicos de los modelos econométricos.

Con este enfoque didáctico recurrente ampliado, se está *instrumentando la estrategia didáctica de reiterar sin saturar los conocimientos básicos* entre los alumnos; por lo que se espera que el lector con esta dinámica gradual, *en cierta* *forma reiterativa y ampliada de la transmisión de los conocimientos*, consolide su cultura econométrica y al final termine bien familiarizado con la teoría y práctica de esta disciplina aplicada al estudio cuantitativo de la economía.

Así, es conveniente decir que la econometría es un producto de la hibridación que se da entre la teoría económica con las matemáticas y la estadística, principalmente; por lo que su marco teórico está constituido por *conceptos* de las tres disciplinas, dentro de los cuales destacan como necesarios para hacer investigación pura y aplicada principalmente los siguientes:

1.- En el ámbito económico es fundamental saber cómo se formula una teoría económica científicamente, es decir, cómo crear apropiadamente los conocimientos que constituyen su cuerpo teórico y que la caracterizan como ciencia. Así, para conformar su arquitectura acostumbro aplicar el método de Oscar Lange (1976). En este contexto conviene que el estudioso de cierto tema económico de su interés inicie su edificación determinando si la teoría económica que formule sobre una variable dependiente, si ésta estará en función de una o de varias variables explicativas y de la naturaleza, tipo y limitaciones de los datos que usará para estructurarla.

2.- En lo que atañe al instrumental matemático que utilizará el investigador para identificar y expresar la relación que guardan entre sí la variable dependiente con la independiente, es importante que sepa la existencia de diversas formas funcionales, sus ecuaciones correspondientes y métodos de cálculo con que se determinan los valores de los coeficientes de las variables explicativas, con el fin de que una de esas formas funcionales sea seleccionada apropiadamente para expresar el presunto comportamiento de la variable dependiente en función de la independiente, la cual sustenta la teoría económica que se está estudiando.

Al respecto, es importante conocer para aplicar correctamente conceptos tan elementales como el de número real, numero natural, el significado de una variable, los tipos de variables que existen, la definición de una función y su aplicación correspondiente, la determinación y características de una ecuación, los métodos con qué se pueden obtener los valores de sus coeficientes y con los que ésta se puede transformar para suavizar el comportamiento de la variable explicada y para que sus coeficientes no pierdan sus propiedades de estimación con una muestra los valores de los parámetros de una población o universo estadístico, entre otros.

Con ese propósito es que en este libro se *han incorporado cuatro apéndices: uno matemático, otro de probabilidad, otro de inferencia estadística y otro de series de tiempo* para que el lector fortalezca, actualice y conozca el instrumental básico que requerirá para hacer investigación aplicada en economía usando la econometría como sustento fundamental de la misma.

3.- Los conceptos estadísticos de gran uso en la verificación de una teoría económica que es conveniente dominar son en general los de la *estadística descriptiva* para identificar y cuantificar con familiaridad las características estadísticas de una distribución de datos, es decir los parámetros. Para ello es conveniente recordar o aprender las medidas de tendencia central, de dispersión y de asimetría, así como los conceptos de números índice, mínimamente.

De manera complementaria pero no menos importante, se debe conocer el significado del concepto de estadística inferencial o inductiva, cuyo fundamento es la metodología del *muestreo estadístico* con la que se pueden *estimar los parámetros* de la población o universo estadístico de la variable dependiente en estudio. En este sentido es conveniente recordar que la selección de la muestra es crucial y que ésta se puede seleccionar en forma empírica o probabilística y que dicha selección con apego a ésta última puede ser con reemplazo o sin reemplazo, para lo cual es necesario saber si se extraerá de un universo finito o infinito porque según sea el caso es que se utilizarán las formulas casuísticamente para determinar su tamaño y margen de error. Por la importancia de este último concepto en la precisión de las estimaciones, aquí es necesario establecer y distinguir entre los conceptos de error estándar, el de muestreo y el permitido con la probabilidad que establezca el investigador.

De esa manera en el caso del muestreo probabilístico, el estudioso de la econometría obtiene una visión clara y suficiente puesto que: 1.- está consciente de que tiene a su disposición diversas opciones (muestras del mismo tamaño seleccionadas probabilísticamente) para hacer las *estimaciones* correspondientes de los parámetros de la población con "*estadísticos*" que le permiten hacer análisis de estructura, de predicción y de evaluación de políticas económicas ; 2.- puede reafirmar sus nociones de estadística en lo que se refiere a la inferencia que se puede hacer con los datos de la muestra sobre las características de la población; 3. Está en condiciones de visualizar su utilidad en la determinación de la relación estructural y de predicción en economía, mismos que le sirven para hacer planeación y por consiguiente, también evaluación de políticas económicas.

En este contexto, como el lector puede observar, de manera *implícita* hemos introducido la teoría de la probabilidad cuya importancia ya fue notoria**,** de manera que es fundamental que el estudioso se entere de las peculiaridades de esta *teoría* ya que al relacionarla con la muestra estadística, su fusión constituye un poderoso instrumento para hacer análisis económico puesto que además de lo antes dicho, con su aplicación es posible identificar y tipificar distribuciones probabilísticas (con sus propiedades respectivas) con las que se puede expresar el comportamiento que pueden tener los datos de las variables económicas ahora vistas en su acepción aleatoria o estocástica, al igual que probar la significación estadística (prueba de hipótesis) que tienen los coeficientes de la ecuación usada para representar la teoría económica, es decir, para verificar que la(s) variable(s) independiente(s) explican o no la variable dependiente: la teoría económica de interés para el investigador.

En esta perspectiva es que también es conveniente saber la importancia que tienen la ley de los grandes números y el teorema del límite central porque sirven para fundamentar que los datos de la variable dependiente tienen una distribución que se asemeja a la normal, situación que le permite al investigador suponer que su distribución tiene las características estadísticas de la normal, lo cual conlleva a: 1.- al manejo sustentado estadísticamente de la gran cantidad de datos que conforman la muestra de esa variable; 2.- conocer la bondad de ajuste que tienen los estimadores muestrales a sus parámetros poblacionales, que muchas veces son desconocidos, 3.- así como identificar las propiedades estadísticas que tienen como estimadores.

Con el dominio de los conceptos descritos en los tres incisos anteriores se configura el marco teórico, el acervo necesario de conocimientos mínimos que el estudioso debe saber y usar para hacer análisis e interpretaciones correctas de los datos que se obtienen con EViews 5.

4.- Se familiariza sobre cómo trabajar la modelística econométrica al dominar el espectro conceptual descrito en los tres puntos anteriores; de ello se intuye que el lector requiere de una formación muy sólida para hacer análisis econométrico apropiado de los resultados a que se llega con Eviews 5.

III.2.--Definición de ciencia y de método científico.

La primera es un conjunto de conocimientos sobre un fenómeno en particular, los cuales se caracterizan por haber sido obtenidos en forma ordenada o sistematizada, por ser congruentes entre si y estar demostrados fehacientemente con evidencias claras y concisas. Lo anterior se logra con la aplicación del método científico, que en opinión de Oscar Lange (1976) es un medio para alcanzar determinados fines, en este caso para obtener los conocimientos, y consta de tres procesos secuenciados, los cuales son: a).- La abstracción; b).- La concretización y c).- La verificación de hipótesis de trabajo con muestras representativas de la realidad del fenómeno que se desea estudiar.

Al constatarse las hipótesis de trabajo, sus conocimientos dejan de ser meros supuestos y pasan a ser conocimientos con los cuales se amplía el estado del arte de la ciencia, los cuales por ser nuevos se denominan “conocimientos frontera”.

Al respecto yo considero que esta descripción del método científico es adecuada porque en la vida real el investigador no está en condiciones de conocer toda la realidad o “universo” de datos que constituyen la teoría económica de su interés. Ante esta limitante, éste se ve obligado a empezar analizándolo parcialmente con muestras, es decir, utiliza el proceso de abstracción, y por consiguiente libera parte del fenómeno e inicia su estudio aplicando el método científico en la forma antes descrita, de manera que en esa forma hace ciencia: aporta nuevos conocimientos que obtuvo con un proceso inferencial y de abstracción sobre el fenómeno, dada la limitante de no conocer en forma directa la realidad total del mismo. Conectando este razonamiento con la teoría económica, puede decirse que ésta es el punto de referencia sobre el cual se va inducir el conocimiento de la realidad aplicando los métodos econométricos que se describen y constituyen la esencia de este libro.

III.3.- La ciencia económica

La ciencia económica, sus conocimientos, así se producen y son científicos porque se obtienen como lo sugiere Oscar Lange (1976:94) y cuya evidencia entre otras cosas los muestra como acciones que se repiten (Sweezy, 1987:22) y que por esa connotación fundamentan las características de las leyes de esta ciencia social, cuya ilustración se ejemplifica, digamos con la ley de la oferta y la demanda, la cual se construye con base en la observación repetida de la conductas racional de los agentes económicos: los consumidores y los productores en un mercado determinado, ante las variaciones de los precios del producto que necesita el primero para satisfacer una necesidad y que produce el segundo en determinadas cantidades. En efecto, se constata repetidamente que: al bajar su precio, el oferente invariablemente reduce su producción y el demandante aumenta su demanda; a la inversa, cuando su precio sube, de manera repetida el productor incrementa el volumen del bien que produce y el consumidor, disminuye su interés por adquirirlo.

III.3.1.- Alcance científico de la econometría:

En la actualidad con base en los avances de los últimos años en que ha mejorado la medición de los datos de las variables que integran el modelo, en que han surgido más y poderosos métodos para conformar la asociación, relación causa-efecto que se establece entre las variables que componen la teoría económica, así como en el manejo de muestras grandes con las innovaciones derivadas del uso de las NTIC (Nuevas Tecnologías de la Información y de las Comunicaciones), por mencionar algunos cambios sustantivos, se estima que la econometría se ha consolidado como una disciplina científica que hace factible el análisis e interpretación de los cambios en la oscilante y compleja economía de estos tiempos; en otras palabras, con los resultados a que se llega y que sí describen la realidad económica, aun que sea en forma parcial (muestra), considero que la econometría es una de las formas más completas de aproximar la economía a las ciencias duras, dentro de las que sobresale la física, que como antes se dijo de ella la economía tomó varios conceptos, leyes, axiomas y teoremas para estudiarla como ciencia y para usarla como instrumento para hacer que la economía mejore la calidad de vida del ser humano.

III.4.- La teoría económica.

Como se ha indicado, las teorías son supuestos, conjeturas o hipótesis que a través de la abstracción el investigador se plantea o construye sobre determinado campo del conocimiento, que en esencia constituye el marco teórico de lo que él desea estudiar, y cuya estructuración se hace seleccionando ciertas variables que en su opinión representan y, por consiguiente, explican dicho campo del conocimiento.

Una vez que se dispone de esas hipótesis, enseguida se procede a verificarlas (Lange, op.cit) estadísticamente usando una muestra proveniente del marco muestral o parte de la realidad que el investigador conoce, cuyos "estadísticos" que suelen ser, por ejemplo, la media aritmética, varianza, proporción, coeficientes de las variables independientes, etc., se contrastan con los supuestos valores de los "parámetros" poblacionales, con un cierto nivel de significación estadística (cuya definición se explicará posteriormente cuando se exponga en detalle cómo se plantea y verifica una *hipótesis estadísticamente*), para inferir si se acepta o rechaza la hipótesis o teoría económica previamente formulada. .

Si se verifica mediante la inferencia estadística la teoría de interés para el estudioso, en ese momento se convierten en ciencia, es decir, en conocimientos sustentados estadísticamente, mismos que se obtuvieron *sistemáticamente* y por ello también tienen la característica de ser *congruentes* entre sí: característica que le da el uso del método científico usado para obtenerlos. Ese acervo será verdad hasta en tanto no surjan nuevos conocimientos que emanen del uso de mejores instrumentos de medición de las variables que integran el marco teórico, de otros métodos aplicados para la obtención, análisis e interpretación de los resultados etc.

III.5.- Concepto de modelo económico.

Vista así la generación de conocimientos científicos; si además de la limitante antes descrita, ahora introducimos el concepto de modelo económico, diciendo que éste es un conjunto de variables que se relacionan y estudian con un fin determinado: describir en forma simplificada una realidad económica especifica (que, como dijimos se visualiza en forma incompleta) explicada a través del establecimiento de una *relación* entre variables que el investigador considera que son representativas de la misma. Con un modelo económico se expresa una conducta racional de los agentes económicos, por ejemplo su reacción ante las variaciones de la oferta y la demanda de un bien de interés suyo. Así, dicho en otras palabras, un modelo estructura teorías económicas modelando r*elaciones entre* variables independientes sobre una variable dependiente, dentro de las primeras se pueden mencionar las combinaciones de los factores de la producción y dentro de la segunda, la fabricación de un bien o el ofrecimiento de un servicio en mercados monopolizados o de libre competencia, ya sea local, regional o internacionalmente.

III.5.1.- Tipos de modelos económicos.

Por el número de variables que constituyen el modelo diremos que existen los de **regresión y correlación lineal simple,** que se caracterizan por el uso de una variable independiente como explicativa de las variaciones de una dependiente.

Los de **regresión y correlación lineal múltiple** son aquellos en que se usa más de una variable independiente para describir los cambios en la variable dependiente.

Se aprovecha para decir de manera inicial que se entiende por regresión a la dependencia que tiene una variable Y de una variable independiente X y que la correlación cuantifica la relación o asociación que existe entre ambas.

III.6.- La bondad de ajuste de los valores de los estimadores.

Diremos que si bien es útil la conceptualización e ilustración de un modelo económico con los ejemplos anteriores dentro de la investigación pura, también es conveniente indicar que los resultados obtenidos con determinada metodología, estos no son validados en función de la cobertura total de una realidad económica especifica, sino de acuerdo con la consistencia lógica de la *especificación económica –matemática q*ue se haya hecho del modelo para expresar, y probar empíricamente, con una muestra dicha realidad económica; en otras palabras, sirve para hacer investigación aplicada a través de la relación funcional de causa-efecto que contrasta la evidencia económica con la especificación teórica. En este contexto, implícita está la situación de que estos resultados no sólo se validan en términos de la metodología aplicada sino también de la "bondad de ajuste estadístico" de los valores de los estimadores a los valores de los parámetros del universo estadístico del fenómeno. Concretamente, la bondad de ajuste de los estimadores se fundamenta en sus características estadísticas detectadas y cuantificadas a partir de una referencia limitada: *la muestra* utilizada, que puede ser o no representativa, además de que puedan o no existir errores de medición, etc.

Lo anterior conlleva de manera implícita a la aceptación de que puede haber errores en " las estimaciones " , mismos que además pueden deberse a que, aun cuando se ha tratado de dar una base científica a las ciencias sociales, en particular a la economía, para cuyo efecto se han tomado muchos conceptos de las leyes de la física, se está plenamente consciente de que a diferencia de las ciencias naturales o exactas en que sus datos permiten elaborar modelos deterministas, al grado de que por ejemplo, se puede predecir con precisión un eclipse. Lo anterior no aplica en la economía, que, aparte de no permitir l*a experimentación* como sucede en las ciencias naturales o exactas, la conducta humana, aún cuando es repetitiva, es impredecible, lo que equivale decir que sus datos no son tan precisos como en esas ciencias, de manera que inserta en los datos de las variables económicas por lo general está presente la *incertidumbre* en ellos, situación que incide en la bondad de ajuste de los estimadores. .

III.7.-.Modelos deterministas versus modelos econométricos o estocásticos.

Para enfrentar esta situación en econometría, se manejan una serie de *supuestos* dentro de los cuales está el de que las variables económicas son aleatorias o estocásticas, es decir, sus datos so*n observaciones* y no resultado de un experimento, tal que el análisis de su distribución permite identificarlos con determinadas distribuciones probabilísticas (normal, t de Student, Chi cuadrada, etc) y por consiguiente, al suponerse que tienen sus características estadísticas, se puede apoyar en ellas y hacer con ellos análisis de estructura, de predicción o de evaluación de políticas públicas, principalmente.

II.7.1.- Modelos econométricos

Así, derivado de lo anterior arribamos a la concepción de los modelos econométricos diciendo que son los que describen relaciones de los protagonistas de la economía, según sea la teoría (la variable dependiente) que se elabore en torno a ellos. En este contexto se puede decir que los modelos econométricos tienen una *variación explicada* por los cambios en los valores de las variables independientes que se incluyan en la teoría y una *variación no explicada* (llamada término de error o valor residual cuyos datos constituyen una variable que también se considera que es aleatoria) que comprende todos los efectos que no provienen de las variables independientes y dentro de los que destacan los errores de medición de los valores de las variables, al igual que el resto de variables que en forma implícita constituyen el complemento de la realidad total que no fue considerada en forma abierta en la abstracción (enfoque parcial o incompleto) que hizo el investigador para configurar la modelística que expresa la teoría en torno a esa realidad económica.

Derivado de lo anterior puede decirse que el alcance de los modelos econométricos es que conjugan los planteamientos o relaciones de causa-efecto que establece la teoría económica que se formule y verifica con el instrumental matemático – estadístico, mediante el cual se cuantifican estas relaciones de causa-efecto, facilitando así la constatación o rechazo de la teoría económica y coadyuvando así a comprender la realidad y a visualizar escenarios predecibles factibles de llevar a cabo por los agentes de la economía, a pesar de que implícitamente contienen las imperfecciones o limitaciones antes señaladas.

III.7.1.1.- Modelos econométricos y métodos de estimación

Como es del dominio público, se dice que en toda investigación los objetivos de la misma determinan los métodos para su consecución. En este contexto es que conviene abordar y discernir sobre los métodos que existen para hacer análisis económico y discernir sobre cuáles son los más apropiados.

Con el fin de enfatizar el uso del concepto de abstracción sugerido por Lange (op.cit) en su concepción de las etapas secuenciadas con que se aplica el método científico en la investigación, que fueron descritas en las secciones previas, ahora es conveniente visualizarla como aquella porción de la realidad que concretiza el investigador para utilizarla en la formulación de su teoría y en su momento, modelarla matemáticamente y evidenciarla estadísticamente.

En este sentido es que conviene, una vez elaborada la teoría económica, describirla con la ecuación seleccionada para expresarla, diciendo que ella, la porción (muestra: cuyos "estadísticos" se representan con letras del alfabeto latino), representa a la población o universos estadístico, cuyos parámetros deben de representarse con letras del alfabeto griego. En esa forma matemática el lector percibe o infiere de inmediato la *relación* de dependencia que tiene la variable Y de la(s) variables(s) X, Z, etc. definidas como independientes o causantes de las fluctuaciones en Y.

Es decir, con un enfoque de didáctica secuenciada, se recomienda escribir con letras del alfabeto latino la ecuación de regresión mediante la cual se estimarán con "estadísticos" provenientes de una muestra, los valores de los parámetros poblacionales, para que al lector le quede claro la distinción entre los parámetros de la población y los "estadísticos" muestrales.

En este contexto en ambas ecuaciones, la de la población y la de la muestra, siempre se debe de incluir la variable "término de error" que así se le denomina porque incluye los errores en la medición de las variables Y y X como las variables explicativas no exhibidas en la ecuación que también impactan los cambios en Y. Así, se denotará como U en la ecuación de la población y como e en la de la muestra.

Por lo que se intuye, la notación es algo convencional, es decir, en la práctica se pueden usar los símbolos que se quiera, la única condición que se pone es que exista un manejo sistematizado con objeto de evitar inconsistencias o confusiones entre los lectores que están asimilando los conocimientos con literales determinadas. Es por ello que algunos autores suelen ponerle un "gorro" a las letras latinas en la ecuación de regresión muestral para enfatizar que se refieren a los estimadores de los parámetros poblacionales (que pueden seguir siendo representados con letras griegas).

Derivado de lo anterior podemos agregar que con la ecuación con letras griegas se expresa el *modelo* en tanto que con la de letras latinas a sus estimaciones correspondientes.

Ahora debemos agregar que se considera apropiado decir aquí que para la formulación del modelo se debe aplicar la siguiente:

III.8.-Relación de la variable estocástica o aleatoria con los modelos.

En economía se modela la teoría con variables conocidas y desconocidas (estocásticas o aleatorias) cuyos datos de las primeras son observados y no procedentes de un experimento. Ello significa que en el *análisis de regresión* se utiliza la *dependencia estadística* entre ellas y no la determinística como digamos en la física; en otras palabras se usa variables que tienen distribuciones de probabilidad cuyas características (media y varianza) hacen factible el análisis de estructura, de evaluación de políticas y de predicción.

Para ilustrar lo anterior Gujarati (2003: 23) presenta el ejemplo de la dependencia del producto de una cosecha de la temperatura, la lluvia, el sol y los fertilizantes aplicados durante la temporada de siembra y cultivo del producto. Comenta que dicha dependencia es estadística porque las variables explicativas no le permiten al agrónomo predecir en forma exacta el producto de la cosecha en función de ellas debido, fundamentalmente, a los errores que surgen durante la medición de las mismas. Ello indica que existe cierta variabilidad intrínseca o aleatoria en la variable dependiente (el producto) "que no puede ser explicada en su totalidad" por ellas y por otras variables explicativas, si es que las hubiera. Ello lleva a decir a Carrascal et al (2001) que las relaciones entre las variables económicas no son exactas y por ello es importante conocer el término de error, es decir, la variable que expresa los errores de medición y/o la omisión de otras variables explicativas del fenómeno en estudio, también llamadas "inobservadas" porque no se ven.

III.9.-Relación de la dependencia estadística (regresión) con la causalidad.

No obstante que el análisis anterior induce a pensar que la variable dependiente depende de otras variables, ello no implica que haya una relación de *causalidad* entre ellas. Ello se debe a que el concepto de causalidad debe provenir de "estadísticas externas y, en último término, de una u otra teoría" Gujarati (2003:22), digamos que para aducir causalidad se debe acudir a consideraciones a priori o teóricas". En resumen, una relación estadística no necesariamente implica una relación de causalidad entre las variables dependiente e independiente.

III.10.- Las variables aleatorias o estocásticas como representantes de las variables económicas

La exposición anterior nos lleva a profundizar sobre la naturaleza y función que desempeñan las variables aleatorias. Así, diremos que puesto que una teoría económica establece relaciones entre variables donde el valor de la que funge como explicada *depende* de los valores que tome(n) otra(s) variable(s) calificadas como explicativas o independientes, esta *relación y dependencia se puede cuantificar usando los conceptos de correlación y de regresión*.En este sentido es conveniente señalar que dicha dependencia o relación entre la variable explicada y la(s) explicativa(s) es *estadística* y no determinística. Luego al ser estadística la relación entre las variables ello indica que se está trabajando con variables definidas como aleatorias o estocásticas por ser ellas las que tienen la característica de que la distribución de sus datos puede ajustarse a alguna de las distribuciones probabilísticas teóricas que aprendemos en los cursos de inferencia estadística y cuyas propiedades nos permiten hacer análisis de estructura y de predicción con una base probabilística.

Al respecto Wooldridge (2009: 847) informa que una variable aleatoria es aquella cuyo resultado es *incierto* y Gujarati (2004:23) comenta que la naturaleza estocástica del *comportamiento de los datos* se puede ilustrar con su ejemplo de “centro del blanco”, cuyo aplicación al tema que nos ocupa la describe diciendo que si existe un tablero que tenga dibujado un centro al que se le puedan arrojar dardos varias veces con el propósito de dar siempre en el blanco, esto último constituye lo que se denomina proceso estocástico, mismo que se construye con los valores que va tomando la variable aleatoria cada vez que se lanza un dardo. Como no hay seguridad de dar en el blanco, de antemano sabemos que es *incierto el resultado* de dar en el blanco y por eso es correcta la definición que hace Wooldridge (op. cit.) de variable aleatoria.

Ahora bien, como el arrojar los dardos tienen siempre el objetivo dar en el blanco, ello lamentablemente no siempre se logra dado que en algunas ocasiones se errará al hacerlo. Errar significa que el proceso de arrojar dardos produce errores. Este razonamiento por analogía se aplica en el análisis de regresión para ilustrar que la dependencia de Y, variable dependiente, de X, variable independiente, es estadística, es decir, que es inexacta, que no está exenta de errores. Esto es así porque por ejemplo, al obtener los valores utilizados para expresar la dependencia de las variables Y de X puede suceder que se cometan errores de medición de los mismos, por señalar una causa de la inexactitud. Así, los errores de medición son equivalentes a los errores que se cometen cuando se tira un dardo varias veces y no siempre da en el blanco. De ello se infiere que las variables usadas en el análisis de regresión son aleatorias o estocásticas, cuyo estudio de la distribución de sus datos puede indicar que tiene semejanza con la distribución de los datos de alguna de las distribuciones de probabilidad teóricas y por consiguiente con base en sus características estadísticas (media aritmética y varianza) se puede hacer econometría empírica determinando la dependencia y/o pronóstico de una variable económica de otra(s) previamente seleccionadas con ese propósito; en otras palabras, con el fin de explicar la variable que representa la teoría económica en estudio.

Luego entonces, en esa perspectiva, al ser aleatorias o estocásticas las variables, ello indica que sus valores no son manipulados porque *surgen al azar como valores observados* y por consiguiente no se puede experimentar para la obtención de ellos, como sucede con los valores de las relaciones determinísticas donde si son manipulables al poder experimentarse con ellos, digamos en ciencias como la física. Así, se puede afirmar que la economía no es una ciencia experimental. Con este referente ahora podemos agregar otra característica de las variables económicas, que sus *datos son conocidos u observados* y que al identificar que es parecido su comportamiento o distribución a los valores de cierta distribución probabilística teórica, al correr la regresión con los datos de la muestra usada para conocer la dependencia estadística de Y de X en su universo estadístico, con sus resultados estadísticos se puede verificar dicha dependencia e inclusive pronosticar su valor futuro., con cierta probabilidad de ocurrencia de sus valores proyectados. Al respecto, su cálculo, como se verá más adelante, se basará en la determinación de cierto valor futuro que se "rodea" de los "límites de confianza" que se construyen en torno a los valores estimados de la variable dependiente, para lo cual se requiere de la determinación del valor del "error estándar de la regresión" y de la "desviación normal estandarizada" (Z o t de Student) asociada a cierta probabilidad de cometer error tipo 1: rechazar algo que es cierto.

En este contexto también conviene citar el alcance que le dan los profesores Galindo y Catalán(2003) a los modelos econométricos al decir que éstos expresan la realidad de un fenómeno económico, por lo que "el modelo teórico es una visión simplificada de la realidad, y por lo tanto, tiene una aplicación parcial". Considero que esta conceptualización es congruente en mucho con los juicios establecidos párrafos arriba por los expertos que allí se mencionan, incluido el suscrito.

En este sentido es que se considero interesante también incluir la percepción que tiene el Dr. Loría (2004: xii) sobre la exactitud de los resultados que se producen con los métodos econométricos cuando dice que los expertos que los elaboran "reconocen ampliamente que habrá errores en la predicción dado que el mundo es estocástico, no determinístico. Es decir, en un mundo donde prevalece la incertidumbre en todas las áreas del quehacer humano lo común es que ocurran eventos totalmente impredecibles (como los del 11 de septiembre de 2001". Está visión complementa y *amplía al futuro* el enfoque que hasta ahora prevalecía sobre el uso de las variables aleatorias o estocásticas como representativas de las económicas, cuyo análisis se venía circunscribiendo a la relación estructural entre variables y a la evaluación de políticas, principalmente públicas.

Con este basamento conceptual ahora procedamos a aplicarlo en la economía.

III.11.- Ejemplo sobre cómo construir una teoría económica.

Dado que los libros de econometría prácticamente no abundan ni profundizan sobre la forma cómo se construyen las teorías económicas y menos sobre las que se refieren a variables específicas de la economía mexicana, se juzgó de interés para los estudiantes mostrar algunos ejemplos sobre los procedimientos que se siguen para estructurarlas, sobre todo, sabiendo que el nivel de estudios de licenciatura con que se introducen en la materia de econometría los limita seriamente para hacerlo con propiedad, lo que es preocupante puesto que, como se ha reiterado, hay que saber cómo se construye una teoría económica porque es la base para hacer econometría, en otras palabras, para especificar el modelo econométrico adecuadamente

En ese contexto es que en este texto al iniciar la exposición de los modelos en que se usan los métodos de regresión y correlación simple con ejercicios ilustrativos, todos ellos empiezan con la exposición de la teoría económica que se desea sustentar econométricamente, de manera muy similar a como se ilustra a continuación la siguiente teoría:

III.11-1.- Ejemplo de formulación de la "Teoría de la competitividad empresarial de las empresas exportadoras de aguacate de Michoacán".

Contexto referencial.

La *teoría* de la competitividad es una de las principales *hipótesis* que en el mundo de la economía y de los negocios de la actualidad, se estudia con ahínco en todos los países, en sus organizaciones empresariales, en sus universidades y en sus centros de investigación avanzada, en virtud de que expresa políticas públicas y privadas, estrategias e ideas sobre los presuntos programas que se deben de instrumentar, a nivel macro y micro, para que tanto los países como sus empresas incursionen exitosamente en los mercados globalizados.

III.11.1.1.-De la competitividad

- Los fundamentos de la competitividad provienen del manejo del método del análisis comparativo entre variables que pueden ser cualitativas y/o cuantitativas, referidas ya sea a bienes o a servicios. Luego entonces su definición y aplicación se fundamentará en el análisis comparativo de las ventajas comparativas y competitivas que se identifiquen en las empresas productoras y exportadoras de aguacate de Michoacán.

- Aun cuando la competitividad se origina (Bonales, 2003:56) en tres niveles: a nivel país, a nivel sector y a nivel empresa, nosotros la estudiaremos sólo a nivel de empresa señalando que es un concepto relativo puesto que no todas las empresas tienen los mismos niveles de competencia en los mercados.

- En este sentido es que *la competitividad se definirá* (Sánchez Barajas, 2007: 45) como “La capacidad que tiene una empresa para incursionar, crecer o consolidarse en el mercado”. Dicha capacidad tiene un enfoque sistémico ya que se halla en cada uno de los departamentos o módulos en que esté constituida la empresa, mismos que en un proceso simbiótico le permiten a ésta posicionarse con su producto/servicio en el mercado temporal o permanentemente. Así es que para evaluar las posibilidades de éxito de un producto y/o servicio en una economía abierta, se propone utilizar un conjunto de variables de naturaleza cualitativa y cuantitativa: estáticas algunas de ellas, otras quizás dinámicas, preferentemente para medir la productividad, la rentabilidad y la eficiencia, etc., con las que se pueda aproximar en mayor medida el concepto de competitividad al nivel específico deseado. Ello significa que con base en el conocimiento de la economía de la entidad y de su potencial de crecimiento estratégico, expresado sectorial y regionalmente, *ahora a nivel de empresa* se deberá de aplicar la metodología que el proceso de investigación recomiende para determinar objetivamente su competitividad. Para ello se sugerirá medir por ejemplo, el dinamismo del mercado, la participación del producto y/o servicio en el mismo, posicionamiento en términos de calidad, precio, tecnología, diseño, marca, organización, disponibilidad de materia prima y de mano de obra, calificada y no calificada; etc.

III.11.1.2.-Proceso de la investigación.

Lo anterior traído el caso concreto que nos ocupa se puede desarrollar mediante el siguiente proceso de investigación:

Primero, se especifica el modelo a utilizar, es decir se establecerá por ejemplo la teoría de que el producto: aguacate (Q) de una empresa depende de los insumos de mano de obra (L) y de capital (K). La cual será la hipótesis a verificar (la teoría es una hipótesis) . Para ello se supondrá que Q tiene una relación positiva con L, es decir, que al aumentar L también aumenta Q, por lo que el signo del coeficiente de L será positivo (+); también, que existe una relación directa entre Q y K, en otras palabras, que a medida que aumenta K, Q aumenta, por lo que el coeficiente de K también tendrá signo positivo (+).

Segundo, se hará una investigación documental para identificar la existencia de datos de corte transversal o de series de tiempo, preferentemente aquellos que correspondan a las variables arriba enunciadas.

Tercero, al contarse con la información, se utilizará el método gráfico del diagrama de dispersión para identificar la forma funcional que se utilizará para hacer los análisis de regresión y de correlación, utilizando el método de mínimos cuadrados.

Cuatro, se obtendrá por ejemplo en esta etapa la ecuación de regresión y se analizarán sus coeficientes o estimadores en una primera instancia, para verificar si sus signos corresponden a la relación que se establezca entre las variables dependiente (Q) e independientes (L,K); enseguida, se evaluará el valor que tomen los coeficientes de determinación y de determinación ajustado con objeto de conocer la relación global entre ellas: en este contexto es que también se calcularán los coeficientes de correlación parcial para detectar el grado en que cada una de las variables explicativas influye en el comportamiento de la variable dependiente. Acto seguido, se hará la prueba de significación para cada uno de los parámetros (prueba t), así como la significación global de todos ellos (prueba F) sobre la Y "calculada o estimada".

Quinto, si estadísticamente se demuestra la teoría de que el producto (Q) depende de las variables exógenas seleccionadas (L,K), se estará en condiciones de hacer un *análisis de estructura sobre la competitividad con el rigor científico* que tiene la econometría; en este caso, de visualizar cómo el producto de la empresa puede resultar más atractivo en los mercados, para ello se deberá hacer un estudio profundo sobre las características de los factores mano de obra y capital con el propósito de localizar su combinación óptima para que genere economías de escala, abata costos de producción, de comercialización y distribución, así como para que en su presentación: volumen de las cajas, calibre o tamaño de los aguacates, grado de maduración, vida en anaquel, especie predilecta .etc. resulte competitiva su penetración en los mercados del país y del extranjero.

III.11.1.2.1.- Pasos a seguir para la formulación de la teoría de la competitividad:

Dado que nunca se tendrá toda la información completa y actualizada sobre el fenómeno que el investigador desea estudiar, para hacerlo se seleccionará una muestra de su universo estadístico y se concreta a analizar su información.

Abstracción:

Los resultados de esta investigación preliminar proporcionan los elementos para *abstraerse,* es decir, para empezar a desarrollar la teoría de la competitividad. Esta debe estructurarse con conceptos generales previos que posea y con los hallazgos concretos que deriven de su investigación preliminar; si se observa que estos se repiten (Sweezy, 1987) o aparecen con cierta frecuencia, ellos constituyen las características del fenómeno en estudio, en este caso las características de la competitividad de las empresas.

Con esa información el investigador empieza a formular su teoría económica de la competitividad empresarial, la cual empieza a describir señalando las relaciones o independencia que existe entre los conceptos, cuyo cúmulo viene a constituir lo que se ha dado en llamar el marco teórico. En el caso de las empresas, el investigador empieza a desarrollar la idea de que las empresas son de diferentes tamaños, es decir que tienen diferentes tamaños de planta y por consiguiente de niveles de producción. Ello lo lleva a señalar que son diferentes los montos de inversión entre ellas y en consecuencia, que es distinta la oferta con que concurren al mercado a satisfacer las necesidades de los consumidores. Otros datos importantes que deben incluirse en el marco teórico son por ejemplo: los procesos de fabricación que usan para generar su producción o servicio cuando la empresa no sea industrial, así como el número de trabajadores que agregan valor a las materias primas e insumos que usan en los procesos productivos. De ello puede concluir que la producción está en función de los factores capital y trabajo y que la competitividad con que actúe en el mercado dependerá mucho de estos dos factores de la producción.

Dado que la generación del conocimiento usando el método científico, entre otras cosas, proviene de un proceso de análisis *secuenciado lógicamente,* se intuye que la siguiente información valiosa que incorpora al marco teórico son las características del mercado, es decir, su capacidad, las características de los diferentes oferentes y consumidores, al igual que los precios con que se hacen las transacciones comerciales de su producto o servicio.

Concretización:

Esta información básica constituye la base para empezar a detectar en forma resumida las fortalezas y debilidades de la empresa, así como, a partir de ellas la *identificación* del problema que tienen que resolver para ser competitivas en el mercado. Así, de manera concreta pero sucinta ya está en condiciones de establecer relación entre las variables, es decir, a formular hipótesis cuya verificación le proporcione le proporcione soluciones al problema de la competitividad empresarial.

Planteamiento y verificación de la hipótesis o teoría de la competitividad:

Algo que salta a la vista de los resultados que obtiene el estudioso del proceso de investigación mediante el cual le fue posible integrar el marco teórico que requiere para conocer la competitividad de las empresas, es el hecho de que la producción (Q) está relacionada con los Trabajadores (L) y con la maquinaria y equipo (K) que de manera combinada usa el empresario en determinado tiempo y lugar para producir Q en una situación competitiva de su empresa.

Este interesante planteamiento teórico lo constata haciendo econometría empírica al usar los datos apropiados de la empresa en estudio, mediante la ***verificación estadística*** de esa hipótesis, que suele llamársele nula, para contrastarla con alguna otra sobre la competitividad de la empresa y, a la que suele llamársele hipótesis alternativa.

III.12.- Especificación concreta del modelo econométrico:

Al ya disponerse de la teoría económica, a partir de este apartado es que ya empieza la econometría a usarse como método con el alcance que antes se le atribuyó: vincular variables, determinar el grado de asociación que tienen entre sí, conocer la bondad de ajuste que tienen los valores estimados a los reales u observados, fundamentar estadísticamente si se verifica o no la teoría económica, hacer análisis de estructura entre las variables en estudio, hacer análisis de predicción, evaluar políticas económicas, etc. Para ello es necesario introducir los *conceptos básicos adicionales* a los antes descritos, dado que se utilizan para la construcción de los modelos econométricos

III.12:1.-Conceptos básicos.

Al ser la econometría la disciplina que expresa una teoría económica a través de las matemáticas y de verificarse con métodos estadísticos, es conveniente señalar que la expresión matemática adopta la forma de modelos, es decir, de ecuacionesque conectan variables específicas para describir la realidad de interés para el investigador.

Al respecto, empezaremos reiterando con otras palabras lo antes dicho: *un modelo es una representación simplificada de la realidad expresada a través de símbolos matemáticos* (José Hernández, 1992:17). Cuando el modelo se relaciona con la economía, se habla de modelos económicos, que son representaciones simplificadas de cierto conjunto de relaciones económicas. Dentro de estos destacan los modelos econométricos, que se definen como modelos económicos que contienen las especificaciones necesarias para su aplicación empírica, es decir *los modelos econométricos constituyen el marco dentro del cual se desenvuelven las investigaciones econométricas*.

III.12.1.1.- Metodología para la formulación y uso de los modelos econométricos.

Para la formulación de los modelos metodológicamente *es necesario realizar las siguientes etapas de trabajo:*

1.- Enunciación de la teoría o hipótesis económica.

2.- La especificación del modelo, que es la exposición de la teoría económica con símbolos matemáticos, es decir, la definición del modelo econométrico dirigido a probar la teoría económica.

3. – Estimación: la determinación del valor numérico de los estimadores muestrales de los parámetros poblacionales del modelo.

4.- Verificación: Es la aceptación o el rechazo de la teoría económica mediante el método de pruebas de hipótesis estadísticas.

5.- Predicción: Se evalúan relaciones estructurales y futuros resultados con base en el modelo establecido.

6.- Aplicación: se evalúan las relaciones estructurales que existan entre las variables endógenas y exógenas, en términos de dinamismo y tendencia principalmente.

7.- Utilización del modelo para fines de control, formulación o evaluación de políticas económicas.

Lo anterior explicado detalladamente significa que, por ejemplo:

III.12.1.1. 1.- *Enunciación de la teoría económica*

Considero que sobre este concepto ya se abundó lo suficiente en incisos anteriores, restaría reiterar que la teoría económica es la base para pasar a hacer *econometría empírica*, puesto que una vez definida se pueden aplicar los procedimientos 2, 3, y 4 (investigación pura o teórica) antes descritos, mismos que una vez constatada la teoría económica, enseguida la ecuación de regresión del modelo se usa para desarrollar los puntos 5,6 y 7 (investigación aplicada).

III. 12.1.1.2.- Especificación del modelo

Conviene indicar que Ursicino Carrascal et al (2001) comentan que el proceso de construcción de un modelo econométrico se inicia con la especificación de la relación a estimar y la formulación de un conjunto de *hipótesis*. Como puede observarse están usando el método científico sugerido por Lange (1976) pero yéndose hasta el tercer paso, las hipótesis. A respecto conviene decir que para formular una hipótesis antes se necesita disponer de información para poder plantearla, por lo que son imprescindibles los primeros dos pasos propuestos por Lange y aplicarlos cómo lo hice con la *teoría de la competitividad*ad antes explicada.

En este contexto conviene decir que *una vez enunciado y especificado* el modelo se plantean las hipótesis en relación con los parámetros de las variables explicitas describiendo la dirección (signo) y fuerza (impacto en la variable dependiente) dentro de la ecuación de regresión que se establezca para el modelo. Así por ejemplo, si el investigador establece en un modelo uniecuacional dentro de las n variables explicativas que n_5 no debe tener efecto sobre la variable dependiente o explicada, ésta hipótesis es equivalente a decir que su parámetro (digamos β) tiene una restricción: debe $\beta=0$; si no es así, entonces puede establecer la hipótesis de que $\beta\geq0$; quizás la hipótesis también pueda ser que $\beta\leq1$.

Por otra parte conviene decir que este procedimiento inicial que requiere seleccionar entre distintas alternativas puede incurrir, sin embargo en errores. Por ello es conveniente someter al modelo elaborado y estimado a diversas pruebas estadísticas que permitan comprobar su validez y calidad antes de utilizarlo en el trabajo empírico. De ahí que se diga que un modelo se especifica cuando se definen variables como las siguientes:

a) **Tipos de Variables**:

*Si decimos que una variable e*s la literal que toma valores dentro del "dominio" determinado por la función matemática, entonces decimos que pueden ser:

*Endógenas***:** son las que explican el modelo y cuyo valor se determina dentro del mismo.

*Exógenas***:** son aquellas cuyos valores no los determina el modelo, son independientes, por consiguiente, influyen en el mismo pero no son influenciadas por el resto de las variables del modelo. No están correlacionadas con el término de error del modelo econométrico establecido.

*Retardadas***:** son aquellas cuyos valores corresponden a momentos, años o periodos de tiempo pasado que influyen en las variables endógenas del presente; pueden ser endógenas o exógenas.

*Predeterminadas***:** son variables exógenas y/o endógenas retardadas en un sistema de ecuaciones simultáneas.

Dicotómicas: **(**también conocidas como ficticias, dummy, *binarias* o categóricas) son las variables cualitativas del modelo, expresadas en atributos, genero, raza, idioma, etc, mismos que se representan cuantitativamente con el 0 y con el 1; es 1 cuando tiene el atributo de interés para el investigador (en el caso del sexo, si es el femenino, cuando la persona tiene sexo masculino, ésta se reconoce con 0 y si la persona es mujer, se registra con el 1).

*Aleatori*a: Variable cuyo resultado es incierto pero predecible probabilísticamente. En ese sentido considera Mason et al (2001:192), que la variable aleatoria es la cantidad que resulta de un experimento aleatorio, mismo que debido al azar, puede generar resultados diferentes. Ejemplo, en el experimento consistente en el lanzamiento de un dado, no conocemos sus resultados posibles, que por cierto son diferentes, pero éstos si se pueden predecir usando la probabilidad; así, al ir acompañado cada resultado con su probabilidad de ocurrencia, da vida a la variable aleatoria o probabilística (estocástica). En este caso la variable aleatoria es *discreta* porque sólo toma ciertos valores que son indivisibles (las seis caras del dado).

Aleatoria de Bernoulli o binaria: Variable aleatoria que toma los valores de cero o de uno.

Aleatoria estandarizada: es $Z= Xi\text{-}\mu/ \sigma$, *la cual tiene* $\mu=0$ *y* $\sigma=1$.

Aleatoria discreta: Toma un número finito de valores enteros no negativos.

Aleatoria continúa: variable que asume valores infinitos en un intervalo determinado.

Instrumental: en una ecuación con una variable endógena explicativa, la variable instrumental no aparece en la ecuación, no está relacionada con el error de la ecuación y se correlaciona (parcialmente) con la variable endógena explicativa (Wooldridge, 2009: 848). Salvatore (1993: 154) es más explicito al decir, en principio que los errores o residuos incluyen errores en la medición de las variables; enseguida señala que la utilidad de esta variable instrumental se ilustra en el caso en que existen errores de medición en la variable explicativa, los cuales conducen a estimaciones de parámetros sesgados e inconsistentes, de manera que para obtener estimadores consistentes se reemplaza la variable explicativa con la *variable instrumental, que está altamente correlacionada con la variable original y que es independiente del término de error. Comenta que "la variable instrumental más simple es usualmente la variable explicatoria en cuestión pero rezagada".*

Endógena explicativa: es una variable explicativa en un modelo de regresión múltiple que está correlacionada con el término de error, ya sea en una variable omitida o a un error de medición o a la simultaneidad.

Endógena rezagada: es el valor rezagado de una de las variables endógenas, en un sistema de ecuaciones simultáneas.

Perturbación aleatoria o término de error: Es una variable aleatoria cuya inclusión se debe al hecho de que las relaciones económicas no se cumplen exactamente (porque son "estimaciones").También se dice que expresa los efectos en la variable dependiente de otras variables exógenas que no se exhiben como explicativas en forma explícita en la forma funcional de la ecuación de regresión. Regresando al enfoque inicial diremos que dicha inclusión, que confiere a los modelos el carácter estocástico, se justifica por las siguientes razones no excluyentes:

1.-Resumen de la influencia conjunta de las *otras* variables exógenas, que, por tener poca importancia, no son incluidas en el modelo de forma independiente.

2.- Recogen los errores de medida en la observación de las variables.

3.- Recogen la aleatoriedad inherente al comportamiento humano.

b).- Estimadores de los Parámetros:

Son coeficientes que acompañan a las variables en el modelo. Son magnitudes que permanecen constantes dentro de un fenómeno concreto y que, por esta razón, reciben el nombre de **parámetros estructurales**. En la teoría económica suelen tener **ciertas restricciones para que ésta** se cumpla, por ejemplo con el signo para indicar que se cumple o no cierta relación entre las variables dependiente e independiente o en su valor: digamos, $0 \leq \beta \leq 1$.

c) Ecuaciones.

Significado: una ecuación es un planteamiento que señala que *dos expresiones son iguales:* Cada una de las *expresiones se llama lado o miembr*o y están separadas por el signo de igualdad (=).

Luego entonces son relaciones matemáticas con las que se trata de expresar la forma en que aparecen relacionados las variables y los parámetros que componen el modelo. Matemáticamente se clasifican así:

Lineales: Cuando la forma funcional es lineal.

No lineales: Cuando la ecuación es de cualquier otro tipo.

Desde el punto de vista económico, las ecuaciones se clasifican en:

De comportamiento: Recogen las acciones de los sujetos económicos.

Institucionales: Describen los efectos de orden jurídico, social y de la política económica sobre el fenómeno en estudio.

Estructurales: Describen la composición del fenómeno bajo estudio, es decir es una ecuación emanada de la teoría económica que se establezca.

Identidad: Son las ecuaciones que expresan relaciones contables o identidades entre magnitudes económicas.

En este contexto de las ecuaciones, cabe agregar que se entiende por:

i).- E*fecto causal:* es el cambio ceteris paribus que tiene una variable sobre otra:

ii).- Efecto marginal: es el efecto que en la variable endógena que resulta de un cambio en la variable exógena.

iii).- *Causalidad de Granger:* Explicación reducida de la causa o efecto que provocan los valores pasados de una variable (Xi) sobre los valores futuros de otra (Yi).

iv).- *Caminata aleatoria con tendencia estocástica:* Es la fluctuación que se observa de una variable en el tiempo, jk que tiene un comportamiento errático y por consiguiente no tiene media ni varianza fijas en el periodo de análisis de la variable.

d).- Datos:

Importancia de los datos en la formulación de los modelos econométricos:

En virtud de que los datos que utiliza el economista u hombre de negocios son datos ***observados*** (como también lo hace por ejemplo el meteorólogo) y por consiguiente no provienen de un experimento controlado como sucede en la biología o en la física. Al no poder controlarlos personalmente, como sucede cuando maneja datos experimentales, para la formulación empírica de modelos, él debe conocer y dominar habilidades adicionales a las que tiene para analizar datos experimentales, además de que debe de estar familiarizado con la naturaleza y estructura de los datos con que pretende verificar su teoría económica.

Para constituir la muestra con la que se estiman los parámetros poblacionales y con ellos cuantificar la dependencia estadística de Y de X, se utilizan tres tipos de datos. Los datos de las variables pueden corresponder a valores de una variable en el tiempo: *series de tiempo*, o a valores para diferentes individuos, grupos u objetos en un momento dado, llamados de*corte transversal* y que generalmente provienen de las *muestras* usadas para hacer encuestas, así como a universos con datos del mismo momento en el tiempo.

Así, los de corte transversal: corresponden a una ó más variables y se caracterizan por ser de *la misma fecha ó mismo punto en el tiempo*, como pueden ser los de los censos de población, que se realizan cada diez; de los censos económicos que se levantan cada cinco años y referirse al ingreso digamos de 30 familias.

Para los de series de tiempo, antes es conveniente decir que una serie de tiempo se define como una sucesión de puntos de una variable económica en cierto periodo de tiempo. En este sentido se dice que los datos corresponden a diferentes momentos o puntos de tiempo en que la variable económica registró información. Pueden ser los valores que toma una acción en una semana, los cinco días en que se registraron constituyen una serie temporal.

A riesgo de correr el riesgo de ubicar el tema aquí y no donde otros opiniones podrían sugerirlo, por su conexión con los datos es conveniente decir que los *datos de panel o longitudinales son:.*

iii).- De panel.

También se les conoce como combinados porque pueden ser los datos del ingreso de 30 familias obtenidos en encuestas que se hacen mensualmente. Su connotación específica es que se refieren a las mismas 30 familias cuyos ingresos son los que varían de una a otra encuesta. Son útiles para medir por ejemplo las variaciones de los ingresos en el tiempo de esas 30 familias.

Son pues estos datos los que corresponden a una serie temporal de una *variable* incluida en una base de datos de corte transversal (personas, empresas, ingresos, etc.). Su característica fundamental (Wooldridge, 2009:10) es que se registran los datos de las variables durante un intervalo de tiempo de: dos, cinco, diez años, etc. de dichas variables de corte transversal.

Alcance del uso de los datos de panel:

1.- Al permitir observar las mismas variables en un periodo de tiempo, proporciona la ventaja que es difícil tener o ver simplemente con datos de corte transversal o en una combinaciones de corte de corte transversal;

2.-Al tener varias observaciones (datos) de la variable ello permite *controlar ciertas características* no observadas de esa misma variable (personas, empresas, etc.) con datos de corte transversal, es decir, al usar más de un dato de una variable f*acilita la inferencia causal* en situaciones en las que inferir causalidad sería muy difícil si se contara sólo con un corte transversal (Wooldridge, 2009: 11).

e).- Clasificación o tipo de modelos:

Pueden clasificarse de diversas maneras, las más comunes son:

i). Por el número de ecuaciones, es decir, pueden ser uniecuacionales o multiecuacionales, todo depende del número de ecuaciones con que se constituyan.

ii). Por la forma de sus ecuaciones, pueden ser lineales cuando sus ecuaciones lo son, no lineales si alguna no lo es.

iii). Por la relación entre el número de ecuaciones y el número de variables endógenas, se dice que es completo cuando coincide el número de ambas; es incompleto, cuando difiere dicho número.

iv). Por el momento del tiempo a que se refieren sus valores: son *estáticos o dinámicos*. En el primer caso todas las variables están referidas al mismo instante de tiempo; en el segundo, cuando dentro del modelo se incorpora alguna variable retardada.

III.12.1.1.3.- Estimación de los parámetros de la ecuación que representa al modelo.

La condiciones técnicas para la construcción y operación de modelos.

En el caso del MLS: Modelo Lineal Simple: uniecuacional, son:

III.12.1.1.3.1.- Presentación:

Como señala el profesor José Hernández Alonso (op. cit., 33), la relación económica más sencilla es la que se establece entre dos variables X e Y que existen en el universo o población estadística, mediante una ecuación lineal ESTRUCTURAL como la siguiente.

$$Y_i = \sim + \quad X_i + U_i$$

Donde:

Y_i: variable endógena, explicada, dependiente o regresada;

X_i: variable exógena, explicativa, independiente o regresora;

Ui: perturbación aleatoria o explicación de los efectos que no explica X_i sobre Y;

α y β: son parámetros desconocidos cuyo valor es necesario determinar (estimar) con el método de mínimos cuadrados u otro.

Nota: El modelo así presentado *se llama modelo econométrico porque contiene la variable Ui*, mismo que se diferencia del *modelo económico o determinista*, cuya notación es: $Y_i = \sim + \quad X_i$
Decimos que el modelo determinista supone una relación *exacta* o determínistica entre las variables Yi e Xi, lo cual generalmente no es cierto en economía por las razones expuestas en párrafos anteriores; por ello es que se considera al modelo econométrico como *más objetivo* dado que incorpora relaciones inexactas entre las variables Yi e Xi, además de que U_i expresa al efecto de otras variables exógenas no explicitadas en la forma funcional de la ecuación de regresión.

El *propósito* inicial de la econometría es *estimar*valores estructurales (de α y β) con valores (â y b̂ que llamaremos "estadísticos") que se obtienen con la ecuación: Ŷ= â+b̂X, a partir de una serie de observaciones o datos muestrales de las variables Y_i, X_i con i= 1,2,3,......,n, mediante la solución de un sistema de ecuaciones con el método de ***mínimos cuadrados***, que permite calcular en la muestra las incógnitas del sistema (â y b̂, llamados "estadísticos" que estiman a α y β) , es decir, conocer los valores estimados de los parámetros propuestos por la relación funcional Y= f(X).

III.12.1.1.3.2 Supuestos básicos del método MCO

Dicho método de MCO (Carrascal et al, 2001:76) se fundamenta en ***el cumplimiento de las siguientes hipótesis o supuestos básicos***:

1. Implícitas en la especificación de la ecuación del modelo está la ***linealidad*** de la relación y la constancia de los ***parámetros.***
2. ***No existen relaciones lineales exactas entre las variables explicativas*** o regresoras y estas no son variables aleatorias (desconocidas).
3. ***Existe linealidad exacta entre las variables solo cuando la variable independiente esta elevada a la potencia 1*** (Gujarati, 1990:32) y se excluyen términos como x^2, $\sqrt{x^2}$ entre otros. **De las dos interpretaciones de linealidad, la de los parámetros** es la más importante en la teoría de la regresión y significa que los parámetros están elevados a la primera potencia. Agréguese que se entiende por linealidad hecho de que todo cambio que experimente X_i en una unidad siempre tendrá el mismo efecto sobre Y_i , el cual viene dado por el valor de β , no importa cual haya sido el valor inicial de X_i.
4. Las perturbaciones aleatorias son variables (desconocidas y por eso se llaman aleatorias o estocásticas) independientes e igualmente *distribuidas normalmente* de media cero y cierta varianza.
5. No existe autocorrelación (son independientes) entre las perturbaciones aleatorias. U_i
6. Todas las perturbaciones aleatorias tienen igual varianza, i.e., hay homocedasticidad, entre las U_i .
7. Existe cero covarianza (cov) entre la variable explicativa (X_i) y la variable o perturbación aleatoria Ui, es decir, no están correlacionadas, de manera que su significación en la variable dependiente (Y_i) es separada y aditiva. Cuando X_i e Ui están correlacionadas (positiva o negativamente) es difícil aislar la influencia individual de X_i y de Ui sobre Y_i (Gujarati, 1990:59).

Bien, estos supuestos se verificarán posteriormente en los capítulos finales de este libro, por ahora debemos decir que se cumplen a plenitud.

Continuando con el tema, ahora nos avocaremos a la:

III.12.1.1.3.3- Verificación de la teoría económica por medio de:

III.1.1.3.3.1.-La inferencia estadística.

En virtud de que se trabaja con muestras y de que con sus unidades muestrales es posible obtener estimadores de los parámetros poblacionales, ello se fundamenta en la estadística inferencial cuyas leyes (como la de los grandes números), y teoremas (como el del límite central), le proporcionan a la econometría las bases suficientes para *inferir* que lo que se *verifica* en la muestra es válido en la *población estadística*.

El procedimiento para inferir lo anterior es el siguiente: empecemos por suponer que usando técnicas estadísticas que luego se explicarán en detalle, digamos como el diagrama de dispersión, que se encontró que la relación entre dos variables que componen la teoría económica, se expresa apropiadamente con la forma funcional de la recta, cuya ecuación llamada de regresión de un modelo lineal simple: MLS, es la (1), la cual usa los valores de Y e X contenidos en una *muestra* para calcular sus respectivos "estadísticos" o estimadores de los parámetros poblacionales ~ y , descritos antes en el punto 3.1. La cual es:

Ŷ= â+b̂X+ ei donde: ………..(1)

Ŷ es estimador de Y

â es estimador de ~

b̂ es estimador de

$\hat{e}_1$ es el estimador de Ui

Para ello si se usa el método de mínimos cuadrados ordinarios, MCO; que así se llama porque minimiza $\sum e_i^2$ puesto que e_i=Y_i-$\hat{Y}_i$=Y_i- â- b̂X_i e indica que es un mínimo la suma de las diferencias elevadas al cuadrado, entre los valores reales y estimados. Así, aplicando las condiciones de minimización que se describen con rigor matemático en el Capítulo IV, se deduce el siguiente "sistema de ecuaciones normales":

$$\sum \boldsymbol{y} = \boldsymbol{na} + \boldsymbol{b} \sum \boldsymbol{x} \qquad (1)$$

$$\sum \boldsymbol{xy} = \boldsymbol{a} \sum \boldsymbol{x} + \boldsymbol{b} \sum \boldsymbol{x}^2 \qquad (2)$$

Con cuya solución se obtienen los valores de las incógnitas *â* y *b̂,* mismos que permiten hacer la inferencia mediante la *prueba de hipótesis*, que en caso de verificarse se constata: a).-la relación lineal de Y con X b).- la teoría económica y por consiguiente, c).- la fundamentación de que lo que es válido en la muestra también lo es en la población estadísticamente hablando. Enseguida se verifica la relación lineal así:

Empecemos reflexionando: ¿Cómo sabemos si los estimadores â y b̂ de los parámetros poblacionales α y β son los adecuados o apropiados?; dicho en otras palabras, ¿De qué manera los métodos estadísticos de estimación pueden aplicarse al análisis de regresión para que éste a su vez fundamente la inferencia estadística, es decir, pruebe la teoría económica?

Bien, si recordamos que cada $\hat{Y}_i$ encaja en la naturaleza de los promedios o medias y que se obtiene a partir de los valores de los estimadores â y ƅ̂, entonces diremos que es posible que ningún valor real de Y sea igual al de $\hat{Y}$ (es decir, ningún punto observado en el diagrama de dispersión cae dentro de la línea de tendencia calculada), pero que también los valores de $\hat{Y}$ pueden estar cercanos de los de Y de manera que la suma de las desviaciones elevadas al cuadrado entre ellas dos es un mínimo, ello induce a pensar que se espera que haya *errores* en todas las estimaciones y que es necesario medir o cuantificar éstas e inferir a partir de ellos el grado de confianza que se puede atribuir a los estimadores; la medición se hace con el cálculo de los valores del error estándar de la regresión y luego del error estándar del coeficiente de regresión: ƅ̂ de X, para luego a partir de su valor hacer la prueba de hipótesis con la *que verificaremos si X efectivamente tiene una relación lineal con Y o que la explica bien*, tal que si se verifica esto último entonces decimos que el estimador ƅ̂ es bueno.

Así, como para hacer la prueba de hipótesis se requiere éste último,i.e., calcularemos el error estándar del estimador o coeficiente de regresión, ƅ̂. Designaremos el error estándar del coeficiente de regresión con $\hat{S}_b$,que se calcula a partir de su ***varianza,*** *cuya fórmula es*:

$$S^2_{\hat{b}} \frac{n\sum y^2-\left(\sum y\right)^2-\left[\frac{\left(n\sum xy-\sum x\sum y\right)^2}{n\sum x^2-\left(\sum x\right)^2}\right]}{(n-2)\left\{n\sum x^2-\left(\sum x\right)^2\right\}}$$

Donde grados de libertad es igual a n-k; n es el número de observaciones en la muestra, k es el número de parámetros a estimar.

Para encontrar $\hat{S}_b$, simplemente le sacamos raíz cuadrada a esta cantidad, así:

$$S_{\hat{b}} = \sqrt{S^2_{\hat{b}}}$$

Continuando con el desarrollo del procedimiento para verificar el efecto lineal del estimador **ƅ̂**, recordemos que hemos supuesto que los errores $e = \left(Y - \hat{Y}\right)$son aleatorios e independientes. Ahora haremos un *supuesto adicional*; que están distribuidos normalmente. **Así, si la distribución de los errores es normal, la cantidad** $t = \dfrac{\hat{b} -}{S_{\hat{b}}}$ **sigue la distribución t con (n-2) grados de libertad.**

Como se indicó, el número de grados de libertad es el número de observaciones, en este caso menos dos debido al número de "estimadores o estadísticos" â y ƃ que han sido previamente determinados con los datos de la muestra, lo cual significa en el análisis de regresión simple que el número de observaciones menos el número de incógnitas calculadas, con los datos de la muestra y usados: â y ƃ para obtener cada $\hat{Y}$, **constituyen los grados de libertad, cuyo valor sirve para mejorar la estimación de β con ƃ.**

Así, en esta ilustración **t** es la medida de la diferencia entre el coeficiente empírico (ƃ) y el coeficiente hipotético (β) de la población, tomando en cuenta la variación de los datos de la muestra.

Con esa formulación de t *probaremos la hipótesis sobre* la ***"veracidad de la teoría económica estadísticamente";*** *posteriormente veremos* que también es *útil para establecer límites de confianza y hacer análisis de predicción.*

Así, diremos que *para constatar la teoría económica establecida por el investigador*, el procedimiento usual es *realizar una prueba de hipótesis* (también llamada de significación estadística) con t, es decir, probar la significación estadística del coeficiente poblacional **β**. Para ello se establece la Hipótesis nula (Ho) de que β=0, que no hay relación lineal entre Y e X, así como la hipótesis alternativa (Ha) de que β≠0, que sí hay relación lineal de Y con X. ***Si se constata la Ho, decimos que no hay una relación líneal*** entre X y Y en la población y que *X no explica a Y* porque al ser β=0, entonces claramente se ve en la ecuación de la población Y=α+βX que (0) X=0, lo cual indica que X no tiene influencia en el comportamiento de Y. para probar Ho se escoge un nivel de significación, usualmente es del 5% y se representa con ~. Si ***rechazamos*** la hipótesis nula Ho, decimos que si hay relación lineal de Y con X y que el coeficiente **β** es significativo estadísticamente, es decir, que es *significativamente diferente de 0*. Por otra parte, si ***aceptamos*** la hipótesis nula Ho, entonces β no es significativamente diferente de 0 y existen indicios de que no hay relación lineal entre X y Y en la población, tal que X no explica a Y

III.12.1.1.3.3.2.-*Ejemplo con datos de corte transversal:*

Supóngase que en un momento dado se tienen los datos de 10 familias y que por motivos prácticos establecemos sin mayor formalidad la teoría económica de que sus consumos (Y) dependen de sus ingresos(X), en un determinado mes del año.

Para que el lector aguce su mente *lo primero que debe hacer es constatar de acuerdo con la definición dada que efectivamente son datos de corte transversal; enseguida, para constatar que la teoría está bien expresada numéricamente, debe realizar una inspección a los datos tanto del consumo como del ingreso. Inmediatamente observará que algunos datos tienen signo negativo y que ciertos datos del consumo* son mayores *que los del ingreso (situación atípica), lo cual no concuerda con la teoría clásica del consumo de las familias. De este análisis preliminar el lector detectará además de las incongruencias anteriores, la importancia que tiene buscar y hallar bases de datos apropiadas. Sin embargo, dado que está teoría del consumo se expondrá con mayor rigor científico* más adelante, *a continuación se usará este ejemplo con el fin exclusivo de introducir e ilustrar los temas anteriores, así como para evaluar la congruencia en los resultados obtenidos con un modelo atípico como este.*

Tabla .1 Ejemplo con datos de corte transversal

Y	***X***	***X²***	***XY***	***Y²***	$\hat{Y}$	$e=(Y-\hat{Y})$	e^2
5	10	100	50	25	3.3227	1.6773	2.8132
-4	-4	16	16	16	-6.6674	2.6674	7.1152
-6	4	16	-24	36	-0.9588	-5.0412	25.4141
-2	1	1	-2	4	-3.0995	1.0995	1.2089
5	16	256	80	25	7.6042	-2.6042	6.7821
-15	-10	100	150	225	-10.9489	-4.0511	16.4112
-8	-12	144	96	64	-12.3761	4.3761	19.1502
10	13	169	130	100	5.4635	4.5365	20.5799
-1	5	25	-5	1	-0.2452	-0.7548	0.5698
-10	-6	36	60	100	-8.0946	-1.9054	3.6306
ΣY=-26	ΣX= 17	ΣX^{2}=863	ΣXY=551	ΣY^{2}=596	-26.0000	0.0000	103.6751

Si la ecuación de regresión en la población es **Y=α+βX+Ui** y si usamos una **muestracon datos de corte transversal en un momento dado,** como la del Cuadro anterior en que la teoría económica establece que el consumo de las familias (Y) depende de su ingreso (X), de manera que usamos sus datos para *estimar* con la ecuación de regresión**Ŷ= â y b̂X+ ei** los valores de sus parámetros α y β. Para calcular dichos valores se usa el sistema de ecuaciones simultáneas explicado párrafos arriba. Ahora bien si ***suponemos que Ui es cero****,* entonces las tabulaciones sobre los *cálculos* necesarios para su obtención se describen de la columna tres a la seis. Manualmentese pueden hacer usando cualquiera de los métodos contenidos en el Apéndice matemático A de este libro. Aquí, en este ejemplo concreto la *solución del sistema de ecuaciones simultáneas* se obtiene con el método de Cramer (ilustrado en el Apéndice A), cuyos cálculos específicos son:

$$\hat{a} = \frac{\sum x^2 * \sum y - \sum x * \sum xy}{n * \sum x^2 - \left(\sum x\right)^2} \qquad \hat{b} = \frac{\sum x^2 * \sum y - \sum x * \sum xy}{n * \sum x^2 - \left(\sum x\right)^2}$$

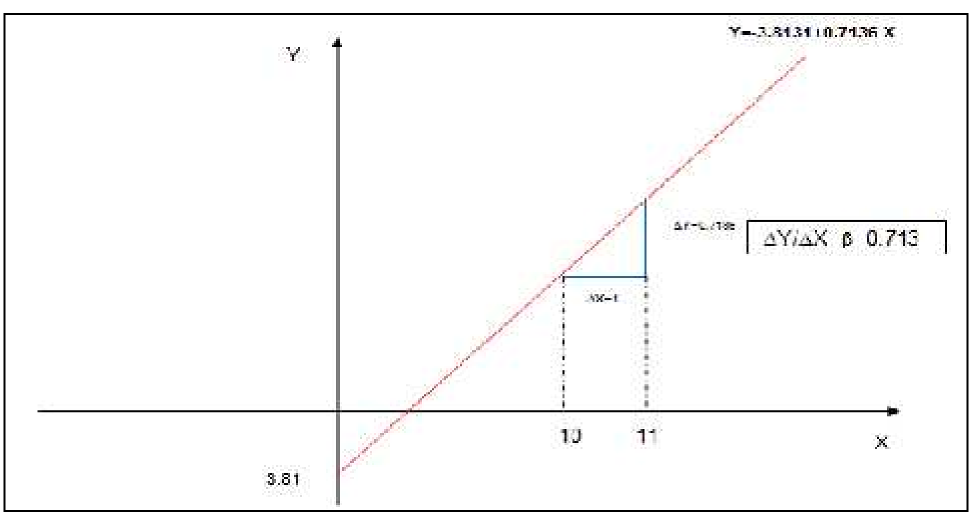
Sustituyendo los valores en cada ecuación nos queda:

$$\hat{a} = \frac{(863)(-26)-(17)(551)}{(10)(863)-(17)^2} = \frac{-31805}{8341} = -3.8131 \quad : \quad \hat{b} = \frac{(10)(551)-(17)(-26)}{(10)(863)-(17)^2} = \frac{5952}{8341} = .7136$$

Luego la ecuación de regresión para estimar Y=α+βX con $\hat{Y} = \hat{a} + \hat{b}x$ es la siguiente:

$$\hat{Y}=-3.8131+0.7136X$$

Esta relación gráficamente así se ve:



Interpretación: si X aumenta en una unidad ($\Delta X=1$), Y lo hace(ΔY) en 0.7136, ya que $(\Delta Y/(\Delta X)=\beta=0.7136$. Por esa razón se dice que a ese cociente en economía también se le llama propensión marginal al consumo, ya que indica que un aumento del ingreso (X) en una unidad, digamos un peso, da lugar a un aumento de 71.36 centavos en el consumo de las familias.

Observaciones: Dadas las incongruencias mencionadas inicialmente, cabe decir que el lector esperaba que el estimador **â** tuviera **signo positivo,** pero no lo tiene; también, que el estimador **$\hat{b}$** observara **signo positivo, que si tiene.**

Ahora bien, si lo que interesa saber es si hay una relación lineal entre X y Y, dado que el signo de $\hat{b}$ resultó positivo, con ello también se puede verificar que el consumo de las diez familias depende de sus ingresos. En este caso los pasos a seguir son: primero calculamos el error estándar del estimador o coeficiente de regresión $\hat{b}$ a partir de la fórmula de su varianza, que es**:**

$$S_{\hat{b}}^2 \frac{n\sum y^2-\left(\sum y\right)^2-\left[\frac{\left(n\sum xy-\sum x\sum y\right)^2}{n\sum x^2-\left(\sum x\right)^2}\right]}{(n-2)\left\{n\sum x^2-\left(\sum x\right)^2\right\}}$$

Donde los grados de libertad son iguales a n-k; **n** es el número de familias (10); **k** es el número de parámetros (2).

Sustituyendo:

$$S_{\hat{b}}^2=\frac{(10)*(596)-(26)^2-\left\{\frac{[(10)*(551)-(17)(-26)]^2}{(10)(863)-(17)^2}\right\}}{(8)\left\{10(863)-(17)^2\right\}}=\frac{1037}{66728}=0.015\cdots luego$$

$$S_{\hat{b}}=\sqrt{S_{\hat{b}}^2}=\sqrt{0.0155}=0.1245$$

Enseguida se establecen la hipótesis nula **H:=β=0 que indica que no hay relación entre Y e X** y la alternativa **Ha:β≠0 que indica que si hay relación entre ellas;** así dado que $n\leq 30$ probamos con t donde

$$t=\frac{\hat{b}-}{S_{\hat{b}}}=\frac{0.7136-0}{0.1245}=5.7317$$

Dado este valor empírico de t, ahora nuestra tarea consiste en escoger la t teórica (o de tablas como le llaman otros autores a t_α) con un nivel de significación dado por el valor que le demos a α, nivel de significación, que expresa la probabilidad de cometer error tipo I, es decir, rechazar algo que es cierto,(el cual es muy distinto al de α aquí usado para expresar la ordenada al origen de la ecuación de regresión de la recta de la población), .al cual probaremos la hipótesis nula **Ho** de que β=0. Supóngase que escogemos el 5% de significación, con n-2=10-2=8 grados de libertad, la tabla indica que $t_\alpha=\pm 2.306$, cuya interpretación es que podemos esperar un valor positivo o negativo de β tan grande como 2.306 si la hipótesis es cierta, es decir, sí el coeficiente de la población es cero. Podemos esperar una diferencia entre cero y el empírico $\hat{b}$, que es el resultado de la selección aleatoria de la muestra, pero esta diferencia no puede ser tan grande como para conducir a valores de t que exceden de 2.306 (positivos o negativos).

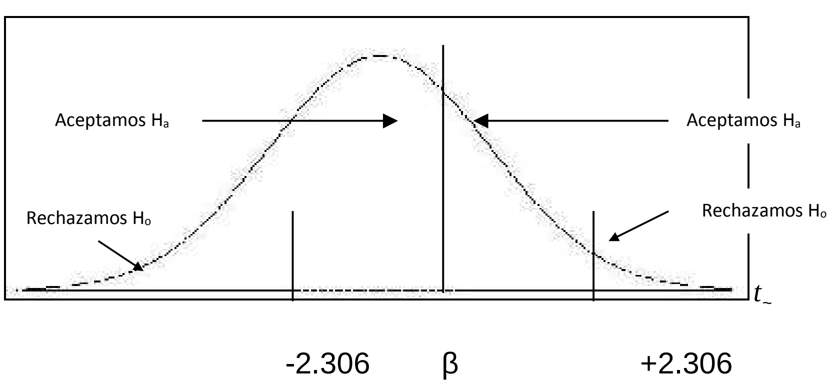
Ahora bien nuestra t empírica (5.7317) es mucho más grande que 2.306, por ello rechazamos la hipótesis nula que β=0 (que no hay relación lineal entre X y Y en la población) y aceptamos la hipótesis alternativa, **Ha** de que β≠0 (si hay relación lineal de Y con X). Decimos que nuestro coeficiente de regresión es significativo estadísticamente a un nivel de significación del 5%. Si la t empírica hubiera sido menor que 2.306, diríamos que el coeficiente de regresión no sería significativo estadísticamente porque La diferencia entre β y cero sería pequeña. En este caso se dice que hay relación lineal porque se rechazó que β=0; en concreto, **aceptamos la Ha que constata la teoría económica de que el consumo de las 10 familias depende de sus ingresos.**

Gráficamente lo anterior se ilustra así:

Si con α=5% y n-2=8 grados de libertad tenemos $t_{\sim} = \pm 2.306$;

llamado punto crítico para aceptar o rechazar H_0, entonces

Gráfica.1 Con datos de corte transversal



Aquí rechazamos Ho porque $t_{\sim} < t_{\hat{b}}$ luego β es diferente de cero, hay una diferencia significativa que no puede atribuirse a la selección aleatoria de la muestra, *sino a los valores de la variable exógena X explicando la variable endógena Y.*

III.12.1.1.3.4-Bandas o intervalos de confianza

Se construyen para "asegurar" con cierta probabilidad que los valores reales u observados se hallen dentro de un intervalo construido en torno a los valores "estimados".

Si decimos que $\hat{Y}$ = **â + b̂X**

$\hat{Y}$: es el valor estimado con los estimadores **â** y **b̂** así como con los cambios que experimente X. Luego la banda de confianza: $\hat{Y} \pm \sigma Z_{\alpha}$ cuando se maneja la población ó muestras grandes; $\hat{Y} \pm \sigma t_{\alpha}$, cuando se manejan muestras pequeñas. Donde $t_{\sim}$ **y** $z_{\sim}$ **:** probabilidad fijada apriorísticamente y, como se recordará, σ es el error estándar de estimación de la regresión.

Uso del error estándar del coeficiente de regresión: $S_{\hat{b}}^2 = \sqrt{S_{\hat{b}}^2}$ **para calcular los límites de confianza que tiene el intervalo de confianza**

Por otra parte para calcular los límites o bandas de confianza, se parte del razonamiento de que algunos parámetros poblacionales se pueden *estimar* calculando el intervalo o banda de confianza en torno al valor del estimador; si el valor hipotético del parámetro esta contenido dentro del intervalo, se acepta la hipótesis; si no, se rechaza. Ahora bien si el nivel de significación es de 5%, ello equivale al error de excluir el valor correcto del parámetro poblacional del intervalo de confianza. Por ello la probabilidad de incluirlo en el intervalo de confianza es de 95%. Decir que $\hat{b}$ es significativo al 5% de nivel de significación, es decir que la probabilidad es de 95%, que los limites de confianza incluya o contengan el verdadero parámetro poblacional.

Los límites de confianza se pueden calcular para un coeficiente de confianza del 95% sustituyendo el valor teórico de $t_{\alpha}=\pm 2.306$ por el valor empírico $t = \frac{\hat{b} -}{S_{\hat{b}}} = \frac{0.7136 - 0}{0.1245} = 5.7317$ resolviendo la ecuación para β obtenemos:

a) 0.7136- β =+2.306(0.1245)=0.2871

β =0.7136-0.2871=0.4265 = Límite inferior

b) 0.7136- β =-2.306(0.1245)=-0.2871

β =0.7136 + 0.2871=1.0007 = Límite superior

Interpretación: Los límites que comprenden el valor del coeficiente verdadero de la población (β) a un nivel de probabilidad del 95% **son**: 0.4265 y 1.0007

Comparación de $S_b \, con \; \pm$ **, donde éste último es el Coeficiente de la regresión, no es del coeficiente de β**

Sabemos que $\pm = \sqrt{\dfrac{\sum\left(\boldsymbol{Y}-\hat{\boldsymbol{Y}}\right)^2}{\boldsymbol{n}-2}} = \sqrt{\dfrac{103.64}{10-2}} = 3.5999$ y que $S_{\hat{b}} = 0.1245$

Se ve que $\pm = 3.5999 > S_{\hat{b}} = 0.1245$ por que este último estima sólo a β, en tanto que el primero es mayor porque estima Y a partir de $\hat{Y}$.

III. 12.1.1.3.5.-Prueba de significación de (r).

Coeficiente de correlación **(r)**:

Se prueba que **Ho**: r=0 y la contrastamos con la hipótesis alternativa de que **Ha**: r es diferente de cero. Si verificamos que r=0, ello indica que X no explica a Y.

Se hace con la fórmula $t_r = \dfrac{r\sqrt{n-2}}{\sqrt{1-R^2}}$ luego trabajando con los datos de la tabla anterior y sabiendo que n=10; r=0.897 y R^2=0.805 tenemos $t_r = \dfrac{0.897\sqrt{8}}{\sqrt{1-0.805}} = \dfrac{0.897(2.37)}{0.442} = 5.74$.

Puesto que con α=5% y 8 grados de libertad la tabla dice que t_α=±2.306 rechazamos la hipótesis que r=0 y decimos que r es significativo con α=5%; que X explica los cambios en Y, *De esta prueba de hipótesis se derivan tres situaciones:*

1.-A medida que r crece es más probable que sea significativamente diferente de cero.

2. Este valor de t=5.74 es igual al de t=5.74 cuando probamos la significación estadística de β porqué la hipótesis de que b=0 implica la hipótesis que r=0.

3.- El signo de r siempre es el mismo que el de β.

III.12.1.1.3.5.1.-Propiedades de los estimadores

La estimación puntual de los parámetros poblacionales α y β se hace por el *método de mínimos cuadrados* que minimiza la suma de los cuadrados de los residuos, $\sum e_i^2$ y que garantiza ciertas *propiedades* estadísticas de los estimadores â y ƅ̂, de los parámetros α y β, que aseguran la confiabilidad del proceso de inferencia (a partir de una muestra se estiman o infieren los valores de los valores de los parámetros). Estas **propiedades de los estimadores** así obtenidos son: que son *lineales, insesgados, óptimos, suficientes, consistentes y eficientes*.

III.12.1.1.3.5.2.- Relación del teorema de Gauss- Markov con las propiedades de los estimadores

Gran parte de estas propiedades descritas de los estimadores fueron demostradas con el teorema de Gauss- Markov, que establece la propiedad del mejor estimador lineal insesgado, cuya cimiente estableció Gauss en 1821 con la invención del método de mínimos cuadrados que sirvió para que Markov en 1900 desarrollara y enriqueciera las propiedades de los estimadores introduciendo el método de varianza mínima con el que pudo calcular los estimadores eficientes o de varianza mínima. De manera que dicho teorema establece y se demuestra que el mejor estimador lineal insesgado es aquel que es lineal, insesgado y tiene varianza mínima.

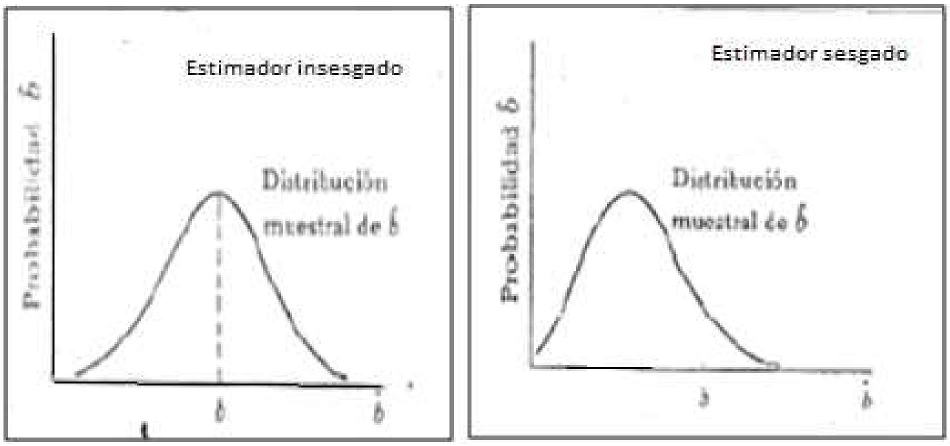
Definición y representación gráfica de los estimadores o "estadísticos muestrales".

Por la importancia del tema conviene profundizar diciendo que su *obtención numérica*en un caso real en este libro se localiza en el Apéndice Estadístico en la sección de "propiedades de los estimadores". Agreguemos que el Profesor Salvatore (1993:107) hizo una buena definición e ilustración gráfica, que es la siguiente:

Insesgado. Una propiedad deseada en un estimador es que este sea insesgado. Un estimador insesgado (b^) es aquel cuyo valor promedio o esperado es igual al valor verdadero del parámetro poblacional (b). <u>Nótese que Salvatore aquí usa el símbolo b para expresar al parámetro poblacional, al cual yo represento con β</u>.

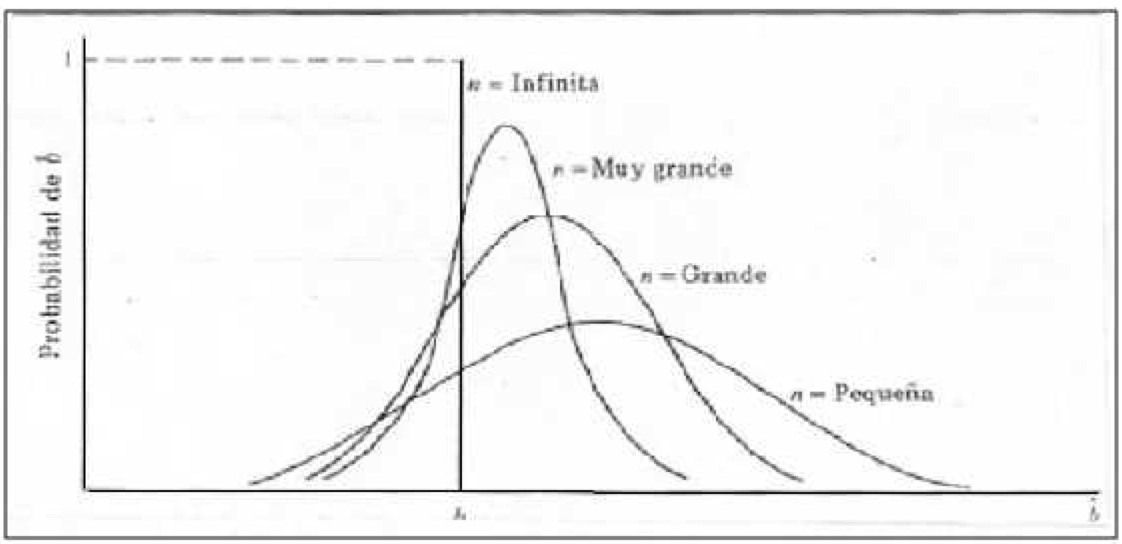
Gráficamente[9]:

Gráfica 2.Estimador insesgado



Consistente. Una segunda propiedad deseada en un estimador es que este sea consistente. Una "estadística" consistente es la que se acerca al valor del parámetro poblacional a medida que aumenta el tamaño de la muestra.

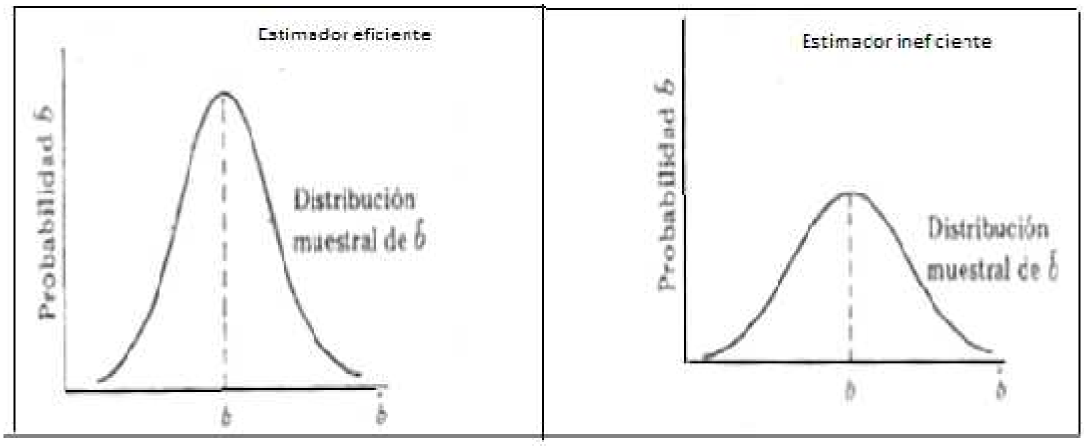
Gráfica 3 Estimador consistente.



Vemos que cuando n crece $\hat{b}$ se acerca a b y que cuando n se acerca al infinito en el límite, la distribución de $\hat{b}$ cae sobre b. Ello es así porque el *estimador Salvatore, 1993:107)* "llega a ser una línea recta con altura (probabilidad) de 1 sobre el valor del parámetro verdadero".

Eficiente. Una "estadística" eficiente es aquella que tiene la varianza mínima entre todos los estimadores disponibles en el marco muestral. En términos del grado, mientras más eficientes es una "estadística" más pequeña es la varianza de su distribución.

Cuadro 3. Estimador eficiente.



Suficiente. Una "estadística" o estimador suficiente es aquel que contiene toda la información disponible de la muestra que usamos para inferir al parámetro poblacional.

***III.12.1.1.3.6*.-Representatividad del modelo descriptivo, su verificación con medidas estadísticas.**

En estadística descriptiva se utilizan dos medidas de dispersión para conocer el "grado de ajuste" o aproximación de los valores estimados a los observados de los parámetros α y β. (capacidad descriptiva) a partir de la ecuación de regresión $\hat{Y}= \hat{a}+\hat{b}X+$ ei, que como sabemos proviene de una muestra. Una de ellas es la varianza de los residuos, medida de dispersión *absoluta*, y otra es el coeficiente de determinación, medida de dispersión *relativa*. Ambas tienen en común que miden la dispersión o alejamiento de los valores estimados con respecto a los valores reales de Y. En su oportunidad se usarán más adelante para conocer la "bondad de ajuste" que tiene el método MCO, *a partir de la cual la ecuación de regresión del Modelo Lineal Simple, MLS, podrá usar para hacer análisis de estructura, de evaluación de políticas y de predicción.*

***III.12.1.1.3.6*.1-Confiabilidad de los estimadores de los parámetros**

La confiabilidad se fundamenta en la estimación puntual de los **parámetros *estructurales*** del modelo: α y β, dada por â y b̂, respectivamente, que provienen de una muestra, los que son dos valores numéricos que puedan provenir de muestras distintas disponibles dentro del *marco muestral* disponible para el investigador, y cuyo *número de muestras que pueden sacarse de éste último* depende de su selección utilizado a partir del muestreo empírico o probabilístico; si se usa este último, de acuerdo con la selección que se haga con o sin reemplazo. Así según la muestra utilizada, es el valor que se obtiene para â y ƀ. Luego entonces para tener una idea de las oscilaciones que pueden producirse en sus valores al pasarse de una muestra u otra, *se calculan las varianzas* de estos dos estimadores: â y ƀ, cuya interpretación es la definida en los cursos de estadística descriptiva e inferencial, i.e., mientras más pequeñas sean, serán más confiables los estimadores muestrales.

***III.12.1.1.3.6.2*.-Importancia del método para estimar**

Derivado de lo anterior surge la importancia de saber escoger el método de estimación .En ese sentido es interesante señalar que *si bien predomina o se usa mucho el de Mínimos Cuadrados Ordinarios, MCO*, ello no significa que sea el único o el mejor, ya que se debe mencionar que también existen el de Máxima Verosimilitud, el de Mínimos Cuadrados Ponderados y el de los Generalizados**,** al igual que el de Mínimos Cuadrados en dos Etapas, así como el de Cocharme .Orcutt, etc. Salta a la vista que debe explicarse por qué se elige cualesquiera de ellos, indicando sus supuestos, en particular la forma funcional, sus propiedades, alcance, y limitaciones en términos de su bondad de ajuste a los valores de los parámetros, etc.

III.12.1.1.4.- Aplicaciones, utilización del modelo.

III.12.1.1.4.1.- Predicción.

Una vez obtenida la ecuación de regresión y de que se ha verificado la bondad estadística de sus estimadores y comprobado que el modelo lineal describe apropiadamente la teoría económica, el investigador está en condiciones de *visualizar el futuro* (hacer planeación) aplicando el modelo uniecuacional o multiecuacional) en la proyección de los valores, digamos por ejemplo los necesarios para *construir escenarios económicos futuros* de su interés, así como para evaluar políticas y hacer análisis de estructura en un periodo de tiempo dado. También, puede usar el modelo para hacer análisis de estructura o para evaluar políticas públicas. He aquí el uso y alcance de los conocimientos básicos de la econometría.

III.12.1.1.5.- ¿Uso de los valores de las variables en sus unidades originales o transformadas?

En esta perspectiva conviene destacar la diferencia que existe en la forma funcional escogida en que se pueden usar variables en *valores originales o en valores transformados*, digamos con logaritmos.Todo indica que en economía convienen usar variables transformadas con logaritmos porque la interpretación de sus coeficientes es atractiva cuando se desea precisar el efecto de las variables independientes sobre la dependiente.

En los modelos en que los coeficientes expresan elasticidades resulta fácil y práctica sus interpretaciones puesto que se refieren a *efectos porcentuales:* En fin como mera sugerencia sobre la selección del método de estimación, es recomendable decir que de conformidad con los objetivos que tenga en mente el investigador, ellos le indicarán cuál de los datos utilizar cuando desee estudiar un fenómeno económico específico.

IV. MODELO LINEAL SIMPLE: MLS.

IV.1. Naturaleza del análisis de regresión y correlación con datos de corte transversal

Congruente con la presentación sugerida del nuevo enfoque didáctico para mejorar la comprensión de los conceptos econométricos, que es consistente con el nivel educativo que poseen los estudiantes de los primeros semestres de la licenciatura en economía, *ya se inició en el Capítulo anterior la exposición de la regresión con datos de corte transversal.*

Cabe reiterar que así como se usa el método didáctico recurrencial para afianzar entre los alumnos el significado delos conceptos que aquí se describen, su aplicación correspondientes se hace de manera ampliada en la sección de ejercicios con NTIC. También es necesario mencionar que estos conceptos se apoyan con el instrumental de matemáticas y estadística que se presentan en los apéndices A y B.

IV.1.1.- ¿Uso de la regresión simple o de la múltiple?

Así, el análisis de regresión se inicia con un modelo lineal simple, MLS, que describe l*a variación promedio* de Y en función de las variaciones de X. Al respecto, se debe comentar que algunos estudiosos de la materia indican que la regresión simple tiene *escaso uso en la econometría aplicada,* afirmación que yo no acepto. Considero que independientemente de su aplicación en econometría empírica, aquí se utiliza para introducir los conceptos básicos de la teoría que subyace entre Y y X puesto que son sencillas las aplicaciones algebraicas que se usan. Pero sobre todo lo describo con detalle porque considero que su uso es imprescindible en todos los países por los micro y pequeños empresarios para trabajar con método sus análisis de mercado, mismos que en México, éstos representan el 98.0% del total registrado en los Censos Económicos que elabora el INEGI quinquenalmente.

La regresión múltiple por supuesto que es importante no sólo porque se ajusta a la realidad económica que suele ser más compleja que la supuesta que describen dos variables. Ejemplo de ello es que se usa para determinar el efecto del gasto público aprobado en cada año fiscal, en los sectores y entidades federativas que comprende el país.

En conclusión yo considero que en la investigación empírica se aplican ampliamente tanto los modelos de regresión lineal simple como múltiple.

IV*.2. Definición del modelo de regresión lineal simple, MLS.*

Se define matemáticamente como

$Y_i = \sim + \quad X_i + U_i$ **……………………………*(1)***

Como se recordará, el significado de las literales ya se dio en el Capítulo III, aquí se reitera y amplia en consonancia con el método didáctico recurrencial.

Así, decimos que Y y X son dos variables correspondientes a una **población** o universo estadístico en que Y se explica en función de X, en otras palabras, la ecuación describe cómo varía Y cuando la hace X.

Ello tomando en cuenta lo dicho en párrafos arriba, considera aspectos como los siguientes: i). que la relación entre Y y X no es exacta; ii). Que por ejemplo a través de diagramas de dispersión aplicados previamente se ha resuelto el problema de expresar matemáticamente su relación con una forma funcional apropiada y iii).- que dicha relación es *ceteris paribus* (es decir, cuando las otras variables no explicitadas en la ecuación no tienen efectos sobre Y)

A la ecuación (1) se le conoce como *modelo de regresión lineal de dos variables*, también llamado por esa razón *modelo de regresión lineal bivariado.*

IV.3.- Significado de las literales de la ecuación (1).

A la literal Y se le menciona con diferentes nombres que van desde variable dependiente, explicada, de respuesta, predicha, endógena, y regresada. A la literal X se le identifica con nombres como los siguientes: variable independiente, explicativa, de control, predictora, exógena y regresora.

De estos términos diremos que todos son sinónimos entre sí para Y y X, respectivamente; sólo precisaremos que el concepto de *independiente* no tiene el significado que se le da en estadística, sino el del algebra: su valor se determina afuera del modelo.

Como indica Wooldridge (2009: 23) los nombres de “explicada” para Y y de “explicativa “ para X son quizás los que describen mejor la relación entre ambas, en tanto que los nombres de “control” y “respuesta” , su uso es más apropiado en las ciencias experimentales dado que allí X es la variable que manipula el investigador para conocer cambios en Y. También, en esta tesitura, se comenta que los términos “predicha” y “predictor” *conviene usarlos preferentemente en el análisis de predicción que en el de causalidad.*

Por su parte a la literal U, como se indicó previamente, se le denomina *“término de error” o perturbación”* en la relación y, sirve para representar los errores en la medición de Y y X, así como para significar a otras variables explicativas de los cambios en Y, que no están explicitadas como X, motivo por el que también a estas últimas se les llama factores no *observados*. En este contexto es que a U también se le puede atribuir la función de representar variables no observabas o que están implícitas en la ecuación (1).

En esa ecuación se observa que si los valores de los conceptos que representa U permanecen constantes: ceteris paribus, entonces ello indica que el cambio denotado por (Δ) en U es cero, situación que permite decir que X tiene un efecto lineal, constante, en Y expresado por:

$\Delta Y = \beta * \Delta X$ cuando $\Delta U=0$…………(2)

La interpretación de (2) es que la fluctuación en el valor de Y es producida por β multiplicada por la fluctuación en el valor de X, donde β es el coeficiente o parámetro de la pendiente de X, en condiciones ceteris paribus ($\Delta U=0$)….

Por otra parte α se le identifica como la ordenada al origen, como *coeficiente del intercepto o también como término constante.*

Unidades de medida heterogéneas

Ahora bien en abono a la bondad estadística de esta metodología de la regresión, es importante destacar que aquí, por ejemplo, como sucede con los números índice, la relación entre Y y X puede establecerse en *unidades de medida diferentes*, digamos que si Y es el salario en pesos, se puede relacionar con X que representa la educación en años de estudio. También debe decirse que cuando se habla de l*inealidad en los parámetros*, se refiere a que una variación en una unidad en X invariablemente produce una variación contante en Y, expresada como β *X.

IV.4.- -Importancia de la teoría económica: eje rector del desarrollo econométrico.

Por lo que se ha visto, puede decirse que la teoría económica constituye el origen y razón de ser de la modelistica econométrica. Ello indica que la econometría es un medio o método para alcanzar los fines económicos, de manera que una vez aceptado el modelo con base en la bondad de ajuste estadístico de sus estimadores con respecto a los parámetros de la población que constituye una realidad económica, se puede crear para crear escenarios sobre relaciones estructurales entre las variables o usarse para hacer pronósticos sobre la mismas.

Así, puede decirse que al usar el método científico de Lange para producir conocimientos, se está en condiciones de generalizar su aplicación hacia cualquier campo del conocimiento, y que al verificarse las hipótesis (que provienen de la abstracción y de la concretización de la realidad del fenómeno de interés para el investigador), estos pasan a ser parte del acervo científico en determinado campo del conocimiento.

En este contexto es que las hipótesis al ser sinónimos de teorías, en la economía constituyen las teorías del mismo nombre, las cuales como se indicó se comprueban con el instrumental de la estadística inferencial.

Conviene señalar que en precisamente durante el proceso de abstracción y de concretización que surgen las teorías económicas, que establecen *relaciones ex* *ante* entre variables para constituir el cuerpo teórico de las mismas y que, generalmente son de causa-efecto o de dependencia, por ejemplo la dependencia de la variable Y de las cambios que operen en la variable X: Dichas relaciones son lo que constituye en econometría "*la especificación del modelo*".

Por consiguiente, de conformidad con la definición que hemos dado de econometría, con un proceso gradual de explicación de transmisión de los aspectos básicos de esta disciplina, ahora decimos que una vez que ya sabemos cómo elaborar una teoría, el siguiente paso consiste en encontrar la manera de describir matemáticamente la relación causa-efecto de X sobre Y. Para conseguirlo es necesario formular criterios matemáticos de especificación de la teoría económica y para su consecución se recomienda recurrir por ejemplo al uso de *diferentes formas funcionales, cuyas ecuaciones y representaciones gráficas a través de los llamados diagramas de dispersión, inicialmente pueden dar un indicio de la mejor forma de describir dicha relación.*

Mediante un proceso de discriminación se arriba a la forma funcional que el investigador considere que "Ajusta "o explica mejor la distribución de los datos de Y e X, cuya dependencia posteriormente se corroborará con los indicadores de bondad de ajuste, de pruebas de hipótesis, etc. Al respecto, *de manera preliminar se puede decir que la forma funcional seleccionada constituye la expresión matemática del modelo econométrico que se usará para explicar una realidad a través de la modelización de la teoría con las variables Y e X*.

De esa manera se compara la evidencia empírica con la especificación teórica, acción que constituye uno de los objetivos medulares de la econometría.

Abundando sobre la dependencia entre variables, se intuye que ésta se establece durante la formulación de la teoría epistemológicamente, dependencia o relación de causa-efecto que la econometría cuantifica basándose en ciertos supuestos que sustentan el método de estimación que se escoja para este fin.

Dicha cuantificación, es decir, la medición de la dependencia de Y de X tiene como fundamento u origen las acciones racionales que se observa llevan a cabo los agentes económicos, que se definen como aleatorias o estocásticas y que se supone tienen una distribución probabilística, cuyas características estadísticas (media y varianza) sustentan la cuantificación de causa-efecto o dependencia de Y de X.

Como puede observarse, la econometría vincula el instrumental teórico económico con el de índole matemático y estadístico. Esta interacción fundamenta la opinión de que un modelo econométrico es una representación sencilla de una realidad económica bajo estudio por parte del investigador.

De lo anterior se deduce que un modelo econométrico aun cuando trabaja una realidad simplificada con información parcial de la misma, en si es muy útil porque establece la simbiosis entre las proposiciones de la teoría económica con los datos “observados” que suponemos que cuantifican o miden la los conceptos de la teoría a través de una teoría estadística.

Así, mediante la abstracción y concretización de la realidad es que se elabora una teoría económica, la cual constituye el referente básico a partir del que se inducen con una muestra las características de la realidad o población estadística del fenómeno bajo estudio.

Como se ha dicho y puede percibirse, la teoría económica *describe* las acciones de los agentes económicos que tipifican la realidad económica de interés, así como las reglas o decisiones racionales que estos toman en su ámbito de trabajo, de manera que la teoría así formulada trata de ser una unidad coherente y sistematizada, o sea es el marco teórico que respalda o justifica la modelización de las variables que, pretendiendo ajustarla a la realidad económica, suelen establecerse ciertas restricciones matemáticas como por ejemplo, que el estimador tenga determinado signo algebraico o que su valor sea mayor que cero pero menor que 1, etc.

En resumen pues, dada una teoría económica que da las bases para especificar un modelo econométrico, este se convierte en el instrumento adecuado para pretender describir, cuantificar y predecir una realidad determinada.

IV.5.- El análisis de correlación como complemento de la regresión.

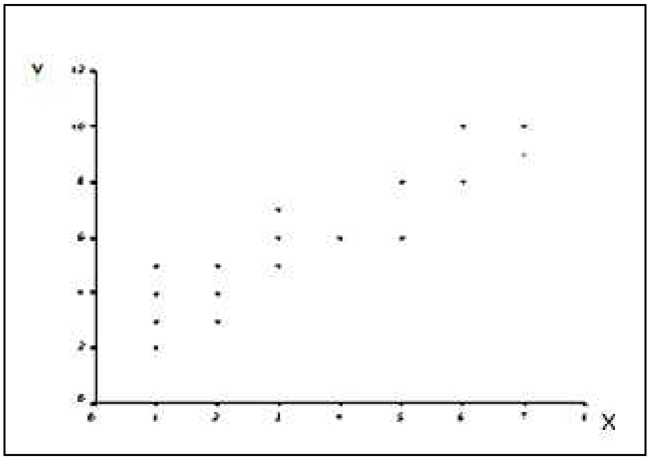
Si sabemos que **Regresión** es la estimación de una variable dependiente, con una ecuación por el método de mínimos cuadrados. La ecuación se denomina de regresión: ***Y = f(X)***.

Ahora bien si la correlación es la determinación del grado de relación que existe entre dos variables por medio del coeficiente de correlación (r), entonces podemos decir que ambas se complementan, en particular si decimos que la correlación también funge como cuantificación del grado de ajuste o cercanía que se logra con los valores de los estimadores a los de los parámetros poblacionales.

IV.5.1. El coeficiente de correlación (r)

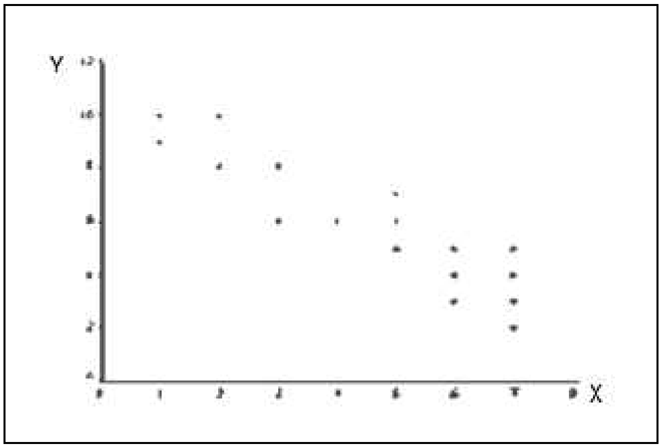
Una vez que hemos obtenido los estimadores, ahora procede la del coeficiente de correlación r**.** Su valor oscila entre -1 y +1 incluido el cero.Decimos que cuando r tiende a +1, ello indica que hay una fuerte relación directa entre Y e X; también decimos que cuando r tiende a -1,que existe una fuerte correlación ó relación, pero inversa entre X y Y. En el primer caso, cuando r tiende a 1 sucede que a medida que X aumenta, también lo hace Y. Ello se ilustra con el siguiente diagrama de dispersión donde las coordenadas de cada uno de los puntos (Y en el eje vertical, X en el eje horizontal) muestran como a medida que X crece también lo hace Y, fundamentando así que r como expresión cuantitativa de su relación tiende a +1.

Gráfica 1 sobre la Correlación positiva.



Por el contrario, en el segundo caso cuando r tiende a -1se indica que las variaciones en X inciden en Y; pero en sentido contrario: a medida que X aumenta Y disminuye. El siguiente diagrama de dispersión muestra como r se aproxima a -1.

Gráfica 2: Correlación negativa.



Relación de los signos de r y del coeficiente β de X.

Suelen ser los mismos es decir, cuando el coeficiente de la pendiente β de X tiene signo negativo, entonces el valor de r es negativoy viceversa.

Relación y cálculo de r con R^2.

Si definimos R^2 como el "Coeficiente de Determinación" y decimos que sus valores oscilan entre 0 y 1, entonces vemos que tiene relación con r sólo cuando sus valor oscila entre 0 y 1; sin embargo en tanto que r mide en términos relativos la relación que existe entre Y y X, decimos que **R^2 indica qué tanto explican los cambios de X los cambios en Y.**

Prueba de bondad de ajuste

Con base en lo antes explicado, ahora conviene señalar que cuando se habla de "ajuste" ello se refiere a la precisión o *cercanía* que observa tener el valor estimado de "la función" de cada uno de los puntos de la recta de Y estimada obtenidos con MCO, con los valores reales u observados de Y. Su cercanía se mide con el coeficiente de determinación (variación explicada) y su diferencia con el error residual(variación no explicada), cuyas literales que los expresan son: R^2 y U, respectivamente.

Para entender mejor lo antes dicho, ahora agreguemos a lo anterior que en realidad existen tres tipos de variaciones relacionadas entre sí (Salvatore: 1993, 91), las cuales son:

$\sum (Y-\bar{Y})^2 = \sum (\hat{Y}-\bar{Y})^2 + \sum (Y-\hat{Y})^2$

Variación total en Y= Variación explicada en Y +Variación residual en Y

(O suma total de cuadrados)= (O suma de cuadrados de regresión)+(O suma de cuadrados de errores)

STC = SCR + SCE

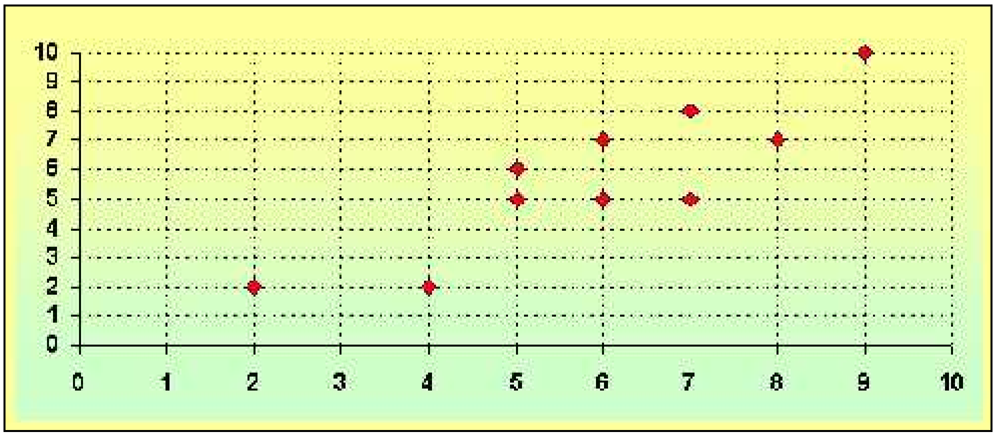
Donde SCE es igual a **e** elevada al cuadrado. Con ello se ve claramente que *en la medida que crece SCR aumenta el valor del coeficiente de correlación y del coeficiente de determinación y, por ende, del ajustado.*

Por su importancia diremos también que SCE mide pues la diferencia entre cada valor de Y real y el que da la línea de regresión. Estas diferencias se originan entre otras causas porque: 1.- hay variables independientes que tienen efectos parciales e irregulares sobre Y, que son omitidos de la relación lineal exacta; 2.- porque quizás existen errores en la medición de Y.

IV.6.-Importancia del diagrama de dispersión

Iniciaremos esta sección diciendo que la regresión y correlación son simples cuando solamente se manejan dos variables(Y,X) y es múltiple cuando se maneja más de dos variables(Y,X,Z,Q) y que en el caso de la primera, la correspondencia gráfica y numérica se puede observar por medio de la forma funcional seleccionada para expresar la relación entre las dos variables, misma que se obtiene mediante el diagrama de dispersión, cuya traza se hace usando el sistema de ejes cartesianos, es decir, se establecen las coordenadas de cada uno de los puntos de Y e X que, al unirlos, dan lugar a una distribución de puntos que es el primer indicio de la posible forma funcional que conviene usar para expresar el modelo lineal de regresión simple, ejemplo, si la figura es la siguiente:

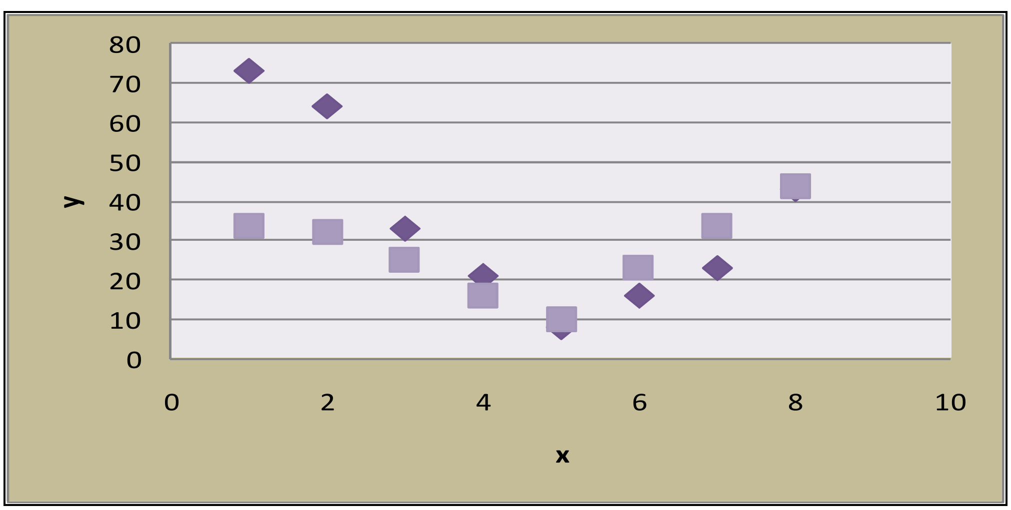
Gráfica 3. Diagrama de Dispersión lineal.



Esto nos índica la distribución que tienen los puntos de las variables. Si tienen la forma anterior, inmediatamente se ve que el ajuste o ecuación a manejar es lineal.

Pero si tiene la siguiente forma, entonces la ecuación a manejar es cuadrática o de segundo grado.

Gráfica 4. Diagrama de dispersión cuadrática.



IV.6.1.-Elección de la forma funcional cuando en el modelo se establece la relación entre más de dos variables.

Hemos visto que el diagrama de dispersión es un instrumento gráfico que nos ayuda a identificar la forma funcional que expresa con propiedad la relación entre dos variables. Al respecto a reserva de explicar más adelante el significado de los indicadores estadísticos, conviene agregar que Gujarati (idem) además propone que se utilicen los siguientes indicadores estadísticos:

Coincide en nuestra exposición al señalar que la elección de una forma funcional particular puede ser relativamente fácil para el caso de dos variables, ya que se pueden graficar las variables y tener así una ligera idea respecto al modelo apropiado; no obstante, advierte que la elección se vuelve mucho más complicada cuando se considera el modelo de regresión múltiple que involucra más de una variable independiente, en cuyo caso sugiere hacer lo siguiente:

1. Es una buena práctica calcular la tasa de cambio (es decir, lapendiente) de la regresada, con respecto a la regresora, así como conocer la elasticidad de la regresada con respecto a la regresora.

2. Los coeficientes del modelo escogido deberán satisfacer determinadas expectativas *a priori.* Por ejemplo, si se está considerando la demanda de automóviles como una función del precio y de otras variables, se deberá esperar un coeficiente negativo para la variable precio.

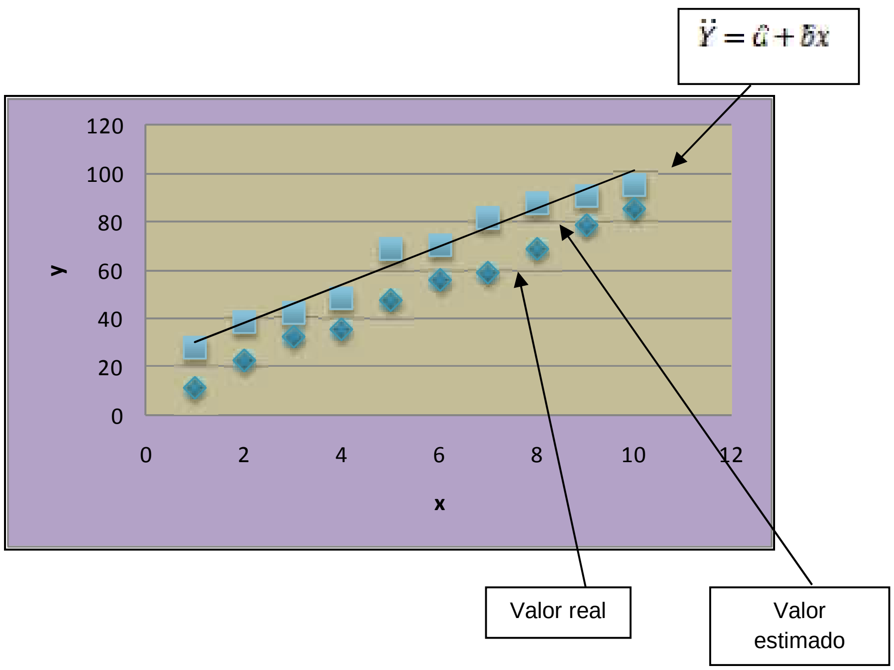
3. Algunas veces tal vez más de un modelo se ajuste razonablemente bien a un determinado conjunto de datos dados.En la curva de Phillips modificada, un modelo lineal y otro recíproco se ajustaron a los datos. En ambos casos, los coeficientes resultaron adecuados para las expectativas previas y fueron estadísticamente significativos. Una gran diferencia fue que el valor R^2del modelo lineal fue mayor que el del modelo recíproco. Por tanto, se podría tener una ligera preferencia por el modelo lineal en comparación con el recíproco. Pero se debe asegurar de que al comparar dos valores de R^2 la variable dependiente (o regresada) de los dos modelos sea la misma; la(s) regresor(as) pueden tomar cualquier forma.

4. En general no se debe sobrevaluar la medición de R^2 en el sentido de creer que mientras más alta sea *r* mejor será el modelo. Si R^2se incrementa conforme se añaden más regresoras al modelo. Lo que reviste mayor importancia es la justificación teórica del modelo elegido, los signos de los coeficientes estimados y su importancia estadística. Si un modelo es bueno bajo estos criterios, entonces quizá resulte aceptable un modelo con una R^2menor.

Al respecto, cualesquiera que sea la *forma funcional seleccionada*, y una vez determinada cuál es la variable dependiente (Y) y cuál es la independiente (X), ésta tiene una “ecuación de regresión” cuyos parámetros (también llamados estimadores) o coeficientes de regresión se obtienen con el método de **denominado de "mínimos cuadrados".** Su nombre se debe a que, como se demuestra enseguida: la suma de las diferencias elevadas al cuadrado entre los valores reales(Y) y los estimados**(Ŷ)** de una muestra es un **mínimo.** Gráficamente se ve así:

En virtud de que la ecuación de regresión es la que aparece en seguida, con la cual se obtienen los valores estimados o "ajustados" (que dan lugar a la recta de regresión) a los valores reales (Cuadros de color azul), la relación entre ambos gráficamente es la siguiente:

Gráfica 5. Identificación de valores reales y estimados.



Así, para hallar los valores de los estimadores (en el caso de la recta, digamos los valores de â y ƀ) se usa el *sistema de ecuaciones normales*, cuya solución mediante cualesquiera de los métodos expuestos en el Apéndice A para encontrar los valores de las incógnitas en un sistema de ecuaciones simultaneas, permite hallar los valores de â y ƀ, mismos que constituyen la parte sustantiva de la ecuación de regresión (en el caso de la recta: Ŷ=â+ƀX) mediante la cual se obtiene un **mínimo con la suma de las diferencias elevadas al cuadrado entre los valores reales y estimados.**

IV.6.2.-- Obtención del sistema de las "ecuaciones normales" (Maddala, 1996:79)

Si dentro de las matemáticas de la optimización definimos Q como un mínimo, tenemos

$Q = \sum \left(\hat{Y}_i - a - bX_i\right)^2$ Hacemos

$2\sum (\hat{Y}_i - a + bX_i)(-1) = 0$

$-2\sum (\hat{Y}_i - \sum \ a + b\sum \ X_i)(-1) = 0$

$-2\sum Y_i + 2\sum a + 2b\sum X_i \ = 0$

Dividimos entre dos y hacemos

$\sum Y_i = na + b\sum X_i$

Si dividimos entre n

$\ddot{Y} = \hat{a} + \hat{b}x$

Si ahora hacemos $\frac{\partial Q}{\partial b} = 0 \Rightarrow 2\sum \left(Y_i - a - bX_i\right)(-X_i)$

$$\sum \left(2Y_i - 2a - 2bX_i\right)(-X_i)$$

$$-2\sum Y_i X_i + 2a\sum X_i + 2b\sum X^2 = 0$$

Dividimos entre 2 y hacemos

$$\sum Y_i X_i = a\sum X_i + b\sum X^2$$

Así el sistema de ecuaciones normales para hallar a y b son:

$\sum Y_i = na + b\sum X_i$ (1)

$\sum X_i Y_i = a\sum X_i + b\sum X_i^2$ (2)

IV.7. Otros Métodos de Estimación

***IV.7.1 Método de Momentos para Obtener los estimadores* â y b̂**

Si con el método de mínimos cuadrados obtuvimos la ecuación $\ddot{Y} = \hat{a} + \hat{b}x$ + ei, entonces $e_i =$ residuos = variable aleatoria, tal que $e_i = Y_i - \hat{Y}_i = Y_i - a - bX_i$, luego las dos "ecuaciones normales" para obtener a y b se obtienen con base en algunas de las hipótesis planteadas; así recordando que $\overline{E}(e_i) = \frac{1}{n}\sum e_i = \sum e_i = 0$ y que $\operatorname{cov}(x_i, e_i) = 0 = \frac{1}{n}\sum X_i e_i = \sum X_i e_i = 0$

Así: i) $\sum e_i = 0$ ó $\sum (Y_i - a - bX_i) = 0$ (1)

ii) $\sum x_i e_i = 0$ ó $\sum X_i (Y_i - a - bX_i) = 0$ (2)

Sabiendo que $\sum a = na$, las ecuaciones (1) y (2) también se pueden escribir de la siguiente manera:

$\sum Y_i = na + b\sum X_i$ ....(1)

$\sum X_i Y_i = a\sum X_i + b\sum X_i^2$ ....(2)

Su solución permite obtener los valores de â y b̂, cuyo valor es idéntico al obtenido con el método de mínimos cuadrados. Si queremos simplificar la obtención de â y b̂ resolvamos simultáneamente las ecuaciones (1) y (2). Así:

$$\hat{b} = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2}$$

$$\hat{a} = \overline{Y}_i - b\overline{X}_i$$

También si hacemos $x_i = X_i - \overline{X}_i$ e $y_i = Y_i - \overline{Y}_i$ tenemos $\hat{b} = \frac{\sum x_i y_i}{\sum x_i^2}$; el parámetro **â** se obtiene igual, es decir $\hat{a} = \overline{Y}_i - b\overline{X}_i$.

k

IV.7.2 Método de Participación de los Residuos para obtener los estimadores â y b̂, así como el Coeficiente de Determinación,(R^2).

Como señala el profesor G.S. Maddala (1996:79) si en las ecuaciones normales sustituimos el valor de a de la ecuación (1) en la ecuación (2), a partir de la ecuación en que $\overline{Y}_i = a + b\overline{X} = a + b\overline{X}$,tenemos que

$\sum Y_i X_i = \sum X_i(\overline{Y} - \hat{b}\overline{X}) + \hat{b}\sum X_i^2$ ....(3)

$$\sum Y_i X_i = n\overline{X}(\overline{Y} - \hat{b}\overline{X}) + \hat{b}\sum X_i^2$$

Con esas referencias, ahora se definen las siguientes **S's**

$$S_{YY} = \sum (Y_i - \overline{Y}_i)^2 = \sum Y_i^2 - n\overline{Y}_i^2$$

$$S_{XY} = \sum (X_i - \overline{X}_i)(Y_i - \overline{Y}_i) = \sum X_i Y_i - n\overline{X}_i \overline{Y}_i$$

así como $S_{XX} = \sum (X_i - \overline{X}_i)^2 = \sum X_i^2 - n\overline{X}_1^2$

La ecuación (3) se puede denotar como $\hat{b}S_{xx} = S_{xy}$ obteniéndose $\hat{b} = \frac{S_{XY}}{S_{XX}}$

así como

$\hat{a} = \hat{Y} - b\hat{X}$

Ahora bien para obtener el coeficiente de determinación (R^2), recordemos que los residuos estimados son $e_i = Y_i - \hat{a} - \hat{b}X_i$, mismos que satisfacen $\sum e_i = 0$ y $\sum X_i e_i = 0$

Si denotamos la suma de cuadrados residuales con RSS donde

$$RSS = \sum\left(Y_i - \hat{a} - \hat{b}X_i\right)^2$$

$$= \sum \left[Y_i - \overline{Y_i} \quad - \hat{b}\left(X_i - \overline{X}_i\right)\right]^2$$

$$= \sum\left(Y_i - \overline{Y}_i\right)^2 + \hat{b}^2 \sum\left(X_i - \overline{X}_i\right)^2 + 2\hat{b}\sum\left(Y_i - \overline{Y}_i\right)\left(X_i - \overline{X}_i\right)$$

$$= S_{YY} + \hat{b}^2 S_{XX} - 2\hat{b}S_{XY}$$

Puesto que $\hat{b} = \frac{S_{XY}}{S_{XX}}$, se tiene $RSS = S_{YY} - \frac{S_{XY}^2}{S_{XX}} = S_{YY} - \hat{b}S_{XY}$

Si expresamos S_{YY} como TSS= Suma de Cuadrados Total, también a $\hat{b}S_{xy}$ con ESS como la suma de cuadrados explicada, decimos que:

TSS = ESS + RSS.
(Variación Total)=(Variación explicada)+ (Variación residual)

Ahora si denotamos con R_{XY}^2=coeficiente de determinación de X en Y = $\frac{ESS}{TSS}$ y $1 - R_{XY}^2 = \frac{RSS}{TSS}$ decimos que R_{XY}^2 explica el efecto de X en Y y que $1 - R_{XY}^2$ es la proporción no explicada por X, o en otras palabras, éste último indica la proporción del efecto de otras variables en Y, distintas a X. Derivado de lo anterior R_{xy}^2 también se obtiene así:

$$R_{XY}^2 = \frac{S_{XY}^2}{S_{XX}S_{YY}} = \frac{bS_{XY}}{S_{YY}}$$

En esta forma ahora se cuenta con nuevas fórmulas para obtener el coeficiente de determinación manualmente.

IV.7.3 Otros Métodos de Estimación.

Existen otros como el de Máxima Verosimilitud, Mínimos Cuadrados de Primera y Segunda Etapa, etc. cuya demostración no se hace en este libro pero se ilustra su aplicación cuando se explica y usa el programa Eviews en el aula de cómputo.

IV.8.- Consideraciones finales para hacer análisis de regresión y correlación simple.

Se dice que el análisis es ***simple*** cuando sólo se estudian dos variables: una dependiente (Y) y otra independiente) y es ***múltiple*** cuando se incluyen en el estudio dos o más variables independientes (X, Z) para explicar los cambios en la variable dependiente (Y).

Debe indicarse que la correlación y la regresión son dos métodos estrechamente vinculados entre si y que se usan para el análisis **de datos muestrales** con los cuales se pretende averiguar cómo se relacionan dos o más variables (Y, X) en una población o universo estadístico.

El análisis de correlación revela si existe o no relación entre Y e X, a través del **Coeficiente de Correlación** (**r**) que resume y cuantifica el grado de asociación entre ellas, en tanto que con el análisis de regresión se determina la dependencia que tiene Y de X, mediante el modelo lineal simple que tiene la ecuación de regresión con la que se **"estiman"** con los datos muestrales los valores de Y a partir de los valores conocidos de X.

De lo anterior se puede inferir que existe una relación de causa-efecto entre las dos variables; sin embargo, debe enfatizarse que con este análisis de regresión únicamente se indica que puede existir una relación matemática entre ellas, dado que la **causalidad** se establece externamente, es decir, en la teoría económica que se formula previamente.

Ahora bien, el método de Mínimos Cuadrados Ordinarios, MCO forma parte de los usados en la **optimización matemática** cuya aplicación se puede hacer a pares de variables (Y,X)o a más variables. Con él se procura hallar una *función* que produce valores estimados de Y (ajuste), que se aproximan bastante a los valores conocidos o reales de ella misma. A la diferencia entre cada uno de los valores estimados y reales se le denomina error o **residuo. La suma de esas diferencias elevadas al cuadrado es un mínimo, de ahí que a este método inventado por Gauss (1821) se le conozca como Mínimos Cuadrados Ordinarios, MCO,** mismo que se usa con mucha frecuencia porque es uno de los mejores para hacer estimaciones de los valores de Y, puesto que son muy parecidos a los valores reales de la misma, tal que sea un mínimo la suma de sus diferencias elevadas al cuadrado.

IV.8.1.- Aspectos básicos a considerar para usar MCO.

Para establecer un modelo que explique Y en función de X se debe de tomar en consideración lo siguiente:

1.- Puesto que entre las variables Y e X nunca existe una relación exacta, ¿Cómo incluir otras variables que también influyan en los cambios de Y?;

2.- ¿De qué forma se puede expresar matemáticamente con propiedad esa relación de dependencia entre Y e X?

3.- ¿Cómo se puede estar seguro de que la dependencia de Y de X sea *ceteris paribus*, es decir, que los valores de otras variables permanezcan contantes?

Para dar respuesta a estas preguntas en el contexto de de la econometría empírica se parte de aspectos eminentemente prácticos como los siguientes:

Se debe de establecer dicha relación con la forma funcional más sencilla que es

$Y=\alpha + \beta X + U$…………………………..(1)

A la cual se le llama *modelo de regresión lineal simple*; como se intuye, por consiguiente, un modelo se define como una relación que se establece entre variables con un fin especifico, en este caso, entre Y e X que representan a una determinada *población o universo estadístico.*

A este modelo representado por una ecuación como la anterior, también se le conoce como de regresión lineal de dos variables o también como de regresión bivariada, dado que indica que existe una relación entre dos variables, la cual es lineal porque los exponentes de α y β están elevados a la primera potencia. Posteriormente se explicará la existencia de modelos univariados, que también expresan la dependencia de Y, pero no de X, sino de sus valores observados en el pasado, así como sus características estadísticas y campos en que se pueden utilizar.

Definición y tipificación de las variables del modelo

A la variable Y se le denomina variable dependiente, variable explicada, de respuesta, predicha o regresada, principalmente; en tanto que a la variable X por ser la que origina los cambios en Y suele llamársele variable independiente, explicativa, de control, predictiva o regresora.

Conviene señalar que que el término "independiente" con el que se ientifica o llama a X, se refirere a que sus valores no se determinan dentro de la ecuación del modelo y que, como se verá más adelante, no tiene el significado "estadístico de independencia entre variables aleatorias" (Wooldridge,2009: 23).

La variable U se conoce como: término de error o perturbación en la relación de Y con X porque representa la aleatoriedad impredecible que introduce la conducta humana (Maddala, 1996:73), así como los errores de medición de Y e X y, también, a las otras variables no observadas que están *implícitas* en la ecuación, que no aparecen como explicativas en la ecuación (1).

Precisiones

Si las variables comprendidas en U se mantienen constantes, de forma que el cambio en U se exprese como $\Delta U=0$, entonces claramente se ve que X tiene *un impacto lineal en Y*, es decir:

$\Delta Y=\beta\Delta X$ si $\Delta U=0$,……………..(2)

De lo anterior se deduce que los cambios que tiene Y provienen de multiplicar la constante β por los cambios que experimente X. Como se recordará, en geometría analítica y por consiguiente aquí, a β se le denomina el *parámetro* (por ser del universo o población estadística) de la *pendiente* en esta relación de causa-efecto de X sobre Y, ceteris paribus: cuando todas las variables no observadas en la ecuación, que están implícitas en U, se mantienen constantes en sus valores.

En este sentido conviene agregar que α se conoce como *parámetro (* por ser del universo o población estadística*) de la ordenada al origen* también llamada *intercepto o término constante (* valor que toma Y cuando X=0). Al respecto, es importante indicar que para algunos estudiosos de la econometría, α no es de gran interés porque no expresa lo fundamental: la determinación de la dependencia que se establece al formular la teoría económica, en otras palabras, para ellos α no tiene una gran importancia en la cuantificación de los efectos que producen los cambios de X en Y.

IV.8.2.- ¿Es o no importante α en las teorías económicas?

Al respecto, yo considero que no se debe de ir a los extremos ya que depende precisamente de qué teoría económica se trate y de la forma como se formule por el investigador, puesto que por ejemplo, en la teoría del consumo (Y) cuando se establece que depende del ingreso (X), siempre será conveniente conocer y cuantificar el consumo básico (α) que se tiene cuando el ingreso sea cero. Por supuesto también existen teorías económicas donde no se considera (α) en cuyo caso está bien ignorarla en el planteamiento de la ecuación de represión. Al respecto, Wooldridge (2009:49) ilustra lo anterior con un buen ejemplo como el siguiente: Si se establece tajantemente que la recaudación Y, depende de los ingresos de los agentes económicos, X, entonces $\alpha=0$, no interesa conocer una recaudación "básica". En este caso para expresar categóricamente la teoría económica en forma matemática la ecuación (1) se transforma en $Y=\beta X+U$.

Reiteración de la importancia de la selección de la forma funcional y su identificación con el diagrama de dispersión.

Su importancia radica en que con su ecuación se determina la ***función***matemática con la que se calculan en la *muestra* los valores *estimados (Ŷ)*de los valores*reales* del universo o población estadística de Y.

¿Pero cuál es la que más aproxima (mejor ajusta) los valores estimados a los reales, es decir dado que existen muchas formas funcionales matemáticas que se pueden usar con ese fin, ¿cuál es la que mejor expresa la *relación* entre Y e X? La respuesta la proporciona el*Diagrama de dispersión*que es una **gráfica**que se obtiene trazando las coordenadas de Y e X en el el sistema de ejes cartesianos. Si al unir los puntos de esa gráfica se observa que sus puntos tienen una tendencia ascendente o descendente, el investigador intuye que la relación entre Y e X se puede describir adecuadamente con la ***forma funcional de la recta;*** por otra parte, si dicha tendencia es similar al comportamiento de la de una curva de segundo grado, se dice que la forma funcional matemática a usar es la de la *parábola*, etc.

Así, si se opta por la forma funcional de la recta, ésta tiene la *ecuación de regresión* (1) cuyos parámetros *poblacionales* α y β al igual que U se *estiman* con los estadísticos *muestrales* â, ƀ, e, respectivamente, que están contenidos en la ecuación de regresión:

Ŷ=â+ƀX+e……………………………………………….……..(3)

Donde Ŷ es el valor estimado de Y con los datos de una muestra determinada; cada uno de los valores de Ŷ se obtiene multiplicando cada uno de los valores de X contenidos en la muestra, cuando ΔU=0. Conviene recordar que cada uno de los puntos de Ŷ son *puntos medios,* dicho en otras palabras, la recta de regresión ajustada a los valores de Y reales, está constituida por *medias aritméticas*

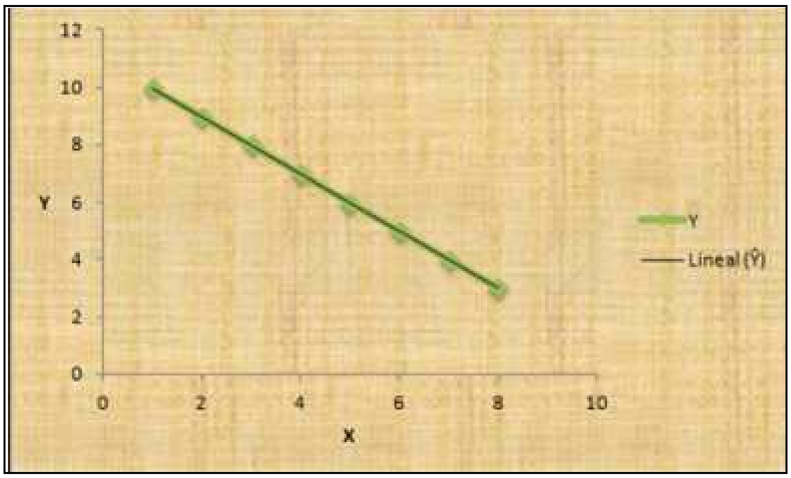
Así, la ecuación de regresión queda de la siguiente forma:

Ŷ= â+ƀX siempre y cuando ΔU=0…………………………….(4)

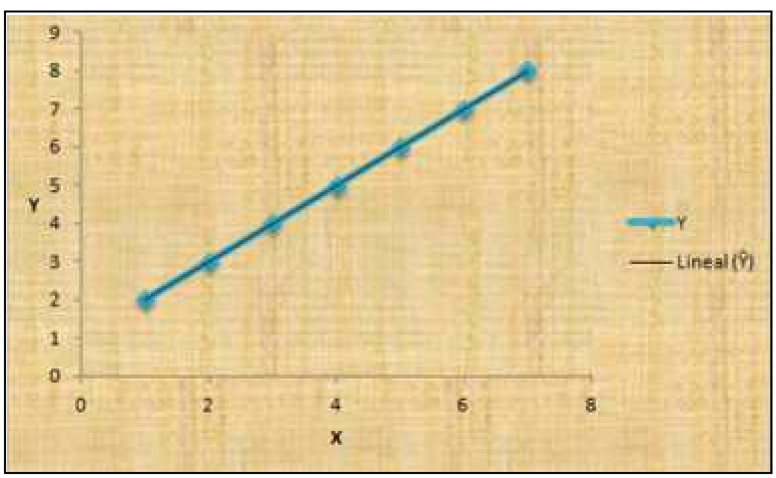
Ahora bien, dado que en la muestra cada uno de los valores de Ŷ es diferente de su correspondiente Y, su diferencia se cuantifica con e= Y- Ŷ= Y– â+ƀX , la cual indica que cuando cierta e es mayor que 0 se refiere al hecho de que Y tiene un valor mayor que su Ŷ correspondiente; en ese caso gráficamente se ve que Y está por encima de su Ŷ (valor promedio) en la curva de la recta de regresión. Por el contrario, cuando cierta e es menor que 0, ello significa que el valor de Y es menor que el de su Ŷ (valor promedio) correspondiente y se sitúa por debajo de ésta última.

IV.8.3.- Situación en que el ajuste de la recta de regresión es perfecto a los valores de Y.

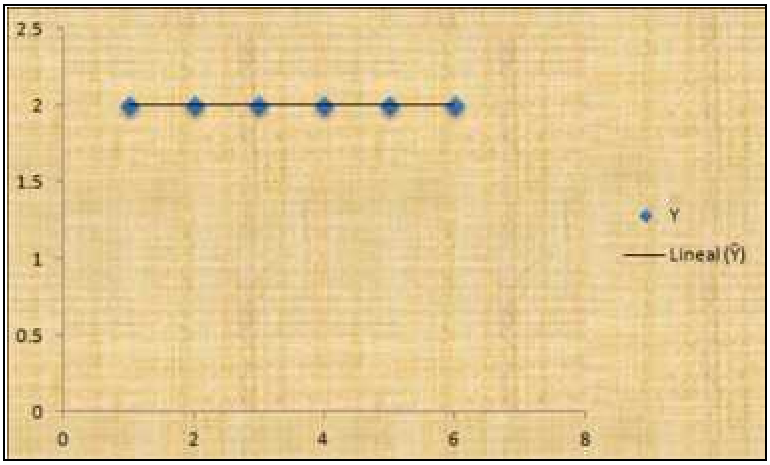
Es importante decir que **idealmente** siempre quisiéramos que e=0, dado que en ese caso Y= Ŷ, lo cual no es posible por la selección aleatoria de la *muestra* que se utilizó para obtener los valores de â y ƀ, por su tamaño, etc. La situación ideal gráficamente se ve en el siguiente diagrama de dispersión donde los puntos de la recta de regresión están sobrepuestos en los puntos reales de Y cuando r=-1 indicando que la asociación o relación de Y con X es perfecta, pero inversa o negativa dado que la pendiente de la recta de regresión es negativa.



Ahora bien cuando r=+1, el siguiente diagrama de dispersión muestra que los puntos de la recta de regresión también están sobrepuestos en los puntos reales de Y indicando que la asociación o relación de Y con X es perfecta, pero directa o positiva, lo cual se constata con la pendiente positiva de la recta de regresión.



Cabe señalar que cuando r=0, el diagrama de dispersión es como el siguiente:



El cual indica que no hay relación o asociación clara entre las variables Y e X.

Continuando con el análisis de la regresión, en este caso:

E(U)=0 ……………....(5)

Por consiguiente U no depende de X de manera que:

La covarianza E(U/X)=E(U)……………………………………………(6)

En la ecuación (6) en que U representa las variables no observables, si su E(U) indica su *"esperanza matemática"* también llamada " *valor promedio esperado",* entonces la ecuación *(6)* revela que su valor promedio es el mismo en todas las U=Y- Ŷ por lo que decimos que U *es una media independiente de X (*Wooldridge, 2009: 25).

IV.8.4.-Aleatoriedad de X y U.

Por otra parte decimos que estas dos variables: X, U son aleatorias porque expresan eventos impredecibles difíciles de prever: Ejemplo de la primera es que X puede provenir de cierta muestra seleccionada aleatoriamente de entre muchas a disposición del investigador. El ejemplo de la aleatoriedad de U se ilustra con cualquiera de las tres fuentes que le dan origen, y que fueron descritas párrafos arriba.

IV.8.5.- Otra forma de detectar la bondad de ajuste.

Ahora bien combinando las ecuaciones (5) y (6) se puede arribar al concepto de "Bondad de Ajuste" en los siguientes términos: Primero, E(U/X)=0 ayuda a visualizar mejor la ecuación (1) ahora en dos componentes: al componente α+βX que representa a E(Y/X) se le denomina *"parte explicada de Y por X"; segundo*, por ende, a U se le conoce como la parte no explicada en Y por X.

Derivado de lo anterior se puede decir que en la medida que la parte "explicada" sea mayor que la "no explicada", ello significa que es *buenala bondad de ajuste* *de los valores* estimados de α y β con â y ɓ, respectivamente. En otras palabras, los valores de Ŷ de la muestra serán próximos a los de Y del universo estadístico, cuya suma del cuadrado de sus diferencias: $\sum(Y-\hat{Y})^2$ es un mínimo obtenido con MCO.

IV.9.-Importancia de la media condicional E(U/X).

El supuesto de media condicional cero: E(U/X)=0 de la ecuación (6) al expresar que U no depende de las variaciones de los valores de X, digamos por ejemplo en el caso de las familias donde Y sea su ahorro y X su ingreso, tal que Y=f(X) indica que si los ingresos de las familias se eligen independientemente de las características de los ahorros, entonces la ecuación (6) *es válida,* ya que el ahorro promedio de las familias no depende de sus ingresos. Por el contrario, si los ahorros crecen con ingresos mayores, entonces el valor promedio o esperado de U variará de acuerdo con el nivel de los ingresos, en cuyo caso la ecuación (6) *no es válida,* es decir, el ahorro de las familias está condicionado por los ingresos que perciban; lo anterior dicho con literales: E(U/X)≠0.

IV.10.- El efecto lineal.

El concepto de la media condicional también ayuda a la obtención e ilustración del *efecto lineal* que tiene β sobre Y, ya que por ejemplo si tomamos el valor esperado de la ecuación (1) supeditado *o "condicionado"* a X suponiendo E(U/X)=0, entonces se obtiene:

E(Y/X)=α+βX…………………………………………………..………….(7)

Ecuación que muestra que la función de regresión poblacional, E(Y/X), "es una función lineal de X, donde la linealidad significa que por cada aumento de una unidad en X *el valor esperado de Y* se modifica en la cantidad β" (Wooldridge, 2009: 26). Al respecto, recuérdese que el valor esperado es el valor promedio de Y que se obtiene en función de las fluctuaciones de X.

Inicio: Una vez que se han obtenido los estimadores (â y ɓ) de de los parámetros poblacionales (α y β) con MCO, de manera que ya se dispone de una recta de regresión que, a su vez, permite "ajustar" los valores estimados (Ŷ) a los reales de (Y) , el siguiente paso es constatar que con esa función matemática *se expresa bien la teoría económica* .Ello se puede hacer en dos formas:

1.- En forma sencilla checando que el signo de la pendiente ɓ sean congruente con la relación establecida entre las variables Y e X durante la formulación de la teoría económica. Si por ejemplo se indicó que había una relación directa, entonces su signo debe ser positivo; por el contrario, si se dijo que era inversa, el signo debe ser negativo.

2.- Formal, en cuyo caso *mediante la evaluación de r: coeficiente de correlación lineal y con el uso de los* valores de los errores estándar de estimadores muestrales de la ecuación de regresión, en una prueba de hipótesis sobre la significación estadística de sus parámetros poblacionales correspondientes, en este caso tanto de α como de β y de r, como se explica a continuación:

Así, de manera secuenciada primero se evalúa el valor de r; si por ejemplo su valor está próximo a -1 o +1, ello induce a pensar que existe una correlación lineal entre los valores muestrales de las variables económicas, Ye X que integran la teoría económica; sin embargo esta r no garantiza que también estén asociadas en la población de donde proceden.

Una buena alternativa a la simple evaluación del valor de r la proporciona la *prueba de hipótesis,* que consiste principalmente en verificar si la pendiente de X en la recta de regresión *poblacional* β es *significativa estadísticamente diferente de cero. Se advierte que los otros parámetros: α y r también se contrastan así; en particular se recomienda que r como β siempre se sometan a este procedimiento de verificación de hipótesis.*

Así, cuando β, α y r resultan ser significativamente diferentes de cero, entonces se dice que en efecto existe *una* correlación lineal *entre Y e X en la población y que se ha constatado la teoría económica que de manera exante estableció dicha relación entre Y e X.* Como *regla tómese nota* de que β y r arrojan los mismos resultados para dar congruencia a la hipótesis planteada, es decir, si por ejemplo β es significativa estadísticamente diferente de cero, también lo es r, y viceversa.

IV.11.- Ampliación del uso de MCO en Regresión y Correlación no Lineal, en la variable explicativa X.

En este caso el comportamiento de Y, como resultado de los cambios en X, no es lineal; sin embargo, dado que los coeficientes son lineales se puede usar MCO para obtener la ecuación de regresión , que en este caso es: $\hat{Y} = \hat{a} + \hat{b}X + \hat{c}X^2$ y para encontrar $\hat{a}$, $\hat{b}$, $\hat{c}$, las ecuaciones lineales serán:

$$\sum y = n\hat{a} + \hat{b}\sum x + \hat{c}\sum x^2 \quad (1)$$

$$\sum xy = \hat{a}\sum x + \hat{b}\sum x^2 + \hat{c}\sum x^3 \quad (2)$$

$$\sum x^2 y = \hat{a}\sum x^2 + \hat{b}\sum x^3 + \hat{c}\sum x^4 \quad (3)$$

Así, si el ejemplo con datos de **corte transversal, es**:

Existen 8 vendedores (Stephen, Shao, 1975) y se desea explicar o estimar sus volúmenes de ventas en función de su experiencia acumulada en un momento de tiempo determinado. Para ello se registran sus ventas (Y), y el número de años de trabajo (X).

Y=f(X)

Observar la siguiente tabla:

Vendedor	Ventas (Y)	Experiencias en años (X)	X^2	X^3	X^4	X*Y	X^2 *Y
A	9	6	36	216	1,296	54	324
B	6	5	25	125	625	30	150
C	4	3	9	27	81	12	36
D	3	1	1	1	1	3	3
E	3	4	16	64	256	12	48
F	5	3	9	27	81	15	45
G	8	6	36	216	1,296	48	288
H	2	2	4	8	16	4	8
Total	**40**	**30**	**136**	**684**	**3,652**	**178**	**902**

Sustituyendo los resultados de la tabla anterior en las ecuaciones tenemos lo siguiente:

$$8\hat{a}+30\hat{b}+136\bar{c}=40$$

$$30\hat{a}+136\hat{b}+684\bar{c}=178$$

$$136\hat{a}+684\hat{b}+365\bar{c}=902$$

Resolviendo el sistema de ecuaciones por el método de Cramer (Apéndice A), el resultado es: $\hat{Y}$ **=3.5914-0.9127X+0.2842X²**.La cual nos permite estimar las ventas **(**$\hat{Y}$**)** en función de los años de experiencia **(X)** de los vendedores.

De esta manera cada $\hat{Y}$ es igual a:

Para A:	$\hat{Y}_1$ **=3.59140.9127(6)+0.2842(6)²=8.3464**
Para B:	$\hat{Y}_2$ **=3.5914-0.9127(5)+0.2842(5)²=6.1329**
Para C:	$\hat{Y}_3$ **=3.5914-0.9127(3)+0.2842(3)²=3.4111**
Para D:	$\hat{Y}_4$**=3.5914-0.9127(1)+0.2842(1)²=2.9629**
Para E:	$\hat{Y}_5$ **=3.5914-0.9127(4)+0.2842(4)²=4.4878**
Para F:	$\hat{Y}_6$**=3.5914-0.9127(3)+0.2842(3)²=3.4111**
Para G:	$\hat{Y}_7$**=3.5914-0.9127(6)+0.2842(6)²=8.3464**
Para H:	$\hat{Y}_8$**=3.5914-0.9127(2)+0.2842(2)²=2.9028**

Si ahora deseamos calcular R^2 y r, para ello decimos que si R^2 =1-$\sum e^2 / \sum (Y-\bar{Y})^2$y quer=$\sqrt{R^2}$, entonces hacemos los siguientes cálculos:

$\hat{y}$	$e=(y-\hat{y})$	$e^2=(y-\hat{y})^2$	$(Y-\bar{Y})$	$(Y-\bar{Y})^2$
8.3464	0.6536	0.4272	4	16
6.1329	-0.1329	0.0177	1	1
3.4111	0.5889	0.3468	-1	1
2.9629	0.0371	0.0014	-2	4
4.4878	-1.4878	2.2135	-2	4
3.4111	1.5889	2.5246	0	0
8.3464	-0.3464	0.1200	3	9
2.9028	-0.9028	0.8150	-3	9
40	0.0	6.4662	0	44

Así, sustituyendo: R^2 =1-6.4662/44=1-0.15=0.85 o 85%.y si decimos que r=√R^2 =√0.85=0.922 o 92.2%, cuya interpretación es la siguiente:

a).- En el caso de r=92.2%, decimos que existe una fuerte asociación o relación de las ventas (Y) de los ocho vendedores con su experiencia en años trabajando(X), la cual se cuantifica por medio de r que tiende a +1.

b).- Cuando R^2 =85%, la interpretación es que su experiencia(X) explica el 85% de sus ventas (Y) y que otras variables (e) el 15% restante.

IV.12.- Aplicaciones con Eview reiterando el marco teórico e ilustrándolo con cálculos sobre las relaciones de variables de la economía mexicana.

A continuación mostraremos el alcance que tiene en la economía la observación que hizo Galdon sobre la relación que existía entre dos variables (Y, X) y la estructuración matemática que hizo su amigo Karl Pearson de dicha asociación haciendo el análisis de correlación e introduciendo el método inventado por Gauss (1821) de Mínimos Cuadrados Ordinarios, MCO, para explicar la regresión: dependencia de la primera (Y) de la segunda(X). Para demostrar lo anterior a continuación se ilustra gráfica y numéricamente la modelización que debe hacerse para calcular y usar los datos de la correlación y de la regresión para estudiar una teoría económica.

Objetivos

1. Conocer su definición, uso y alcance;
2. Conocer y aplicar sus fórmulas para calcularlos;
3. Introducir, trazar e interpretar el concepto de diagrama de dispersión para ampliar el análisis sobre la relación de variables;
4. Presentar la ecuación de la forma funcional que describe mejor la relación entre la variable dependiente (Y) e independiente(s) X,Z,Q,………
5. Saber calcular los coeficientes, “estimadores” o incógnitas de los valores de los parámetros de la ecuación de regresión con MCO, así como saber

interpretar estadística y económicamente dichos valores; en particular se hará énfasis en el *ajuste* de los valores estimados a los observados ;

6. Realizar inferencias sobre los parámetros de la población o universo estadístico que representa la forma funcional, inicialmente de la recta de regresión para después sofisticar el estudio;
7. Predecir e interpretar intervalos de confianza e intervalos de predicción para la variable dependiente;
8. Realizar una prueba de hipótesis para verificar si con este modelo se constata la teoría económica establecida previamente.

IV.12.1.- Ejemplos utilizando EViews 5.

La ilustración de la regresión simple se realizará mediante el uso de EViews 5 utilizando datos actuales de la economía mexicana.

IV.12.1.1.- Ejemplo 1: Teoría keynesiana del consumo.

Planteamiento de la teoría económica

Partiendo de la teoría keynesiana del consumo se utilizarán datos de la economía mexicana del consumo privado (de las familias) y del Producto Interno Bruto (PIB), desde su acepción como ingreso[9], para el periodo comprendido entre el primer trimestre de 1980 y el cuarto trimestre de 2010, ambas en miles de pesos a precios constantes de 2003.

Así, el modelo matemático que se plantea tiene la siguiente estructura:

Consumo = f (PIB)

Donde el Consumo está en función del ingreso, en este caso del PIB.

Mientras que el modelo econométrico es el siguiente:

$$Consumo = \hat{a} + \hat{b}PIB + \hat{e}$$

La relación teórica entre el PIB y el Consumo es directa, es decir, si el PIB (el ingreso) crece se espera que el Consumo también lo haga y viceversa, si el PIB disminuye el Consumo también lo hará.

De acuerdo con la teoría $\hat{a}$ representa el consumo autónomo[10], mismo que debe ser mayor a cero, pues la sociedad debe poseer un consumo mínimo positivo independientemente de su ingreso, por tanto:

$$\hat{a} > 0$$

---

[9] El PIB puede ser calculado o analizado desde tres concepciones: 1) Método del gasto; 2) Método del ingreso; y 3) Método del Valor Agregado.

[10] En matemáticas $\hat{a}$ es la ordenada al origen.

Por su parte $\hat{b}$ representa la Propensión Marginal a Consumir (PMgC), misma que mide cuanto se incrementa el consumo cuando se incrementa en una unidad el ingreso[11], por tanto, el valor que asume $\hat{b}$ se encuentra entre 0 y 1:

$$0 \leq \hat{b} \leq 1$$

En el análisis de la teoría keynesiana del consumo la diferencia 1- $\hat{b}$ representa la Propensión Marginal a Ahorrar (PMgS):

$$PMgS = 1 - \hat{b}$$

Así, si $\hat{b} = 1$, el individuo o las familias gastan completamente la unidad adicional de ingreso que obtienen, por lo que la PMgS es igual a cero; si $\hat{b} = 0$, el individuo o las familias no gastan la unidad adicional de ingreso y se presupone que entonces lo ahorran en su totalidad, por lo que la PMgS es igual a 1; cuando $0 < \hat{b} < 1$, entonces existe cierta PMgS que es igual a 1- $\hat{b}$.

Ahora bien, en cuanto al software, EViews tiene una estructura tipo Windows, la cual contiene una barra de menú, una barra de comandos, el área de resultados o de trabajo y una barra de estado de la aplicación (ver Cuadro 1.1).

Ejemplo 1: Teoría Keynesiana del Consumo.

Cuadro 1.1. Estructura de EViews.

<u>Creación del archivo de datos.</u>

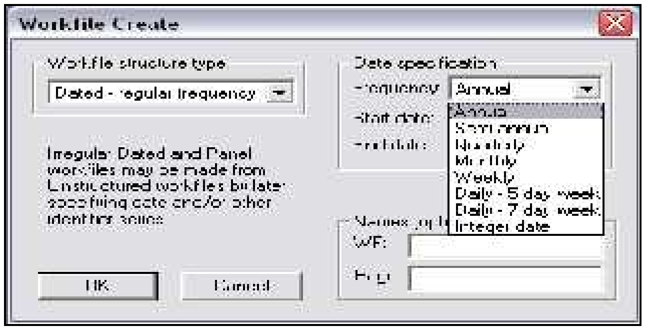
[11] En matemáticas $\hat{b}$ es la pendiente.

Para generar o crear un nuevo archivo de trabajo, desde la barra de menú vamos a File/ New/ Workfile con lo cual se desplegará una ventana en la cual deberá especificarse la frecuencia de los datos (Date specification) e inclusive, de manera opcional, el nombre del archivo de trabajo (Workfile) y de la página de éste (Names (optional)). Asimismo, también puede obtenerse dicha ventana escribiendo en la barra de comandos la palabra "new" seguido de la tecla enter. Cabe mencionar que en la opción de tipo de estructura del archivo de trabajo (Workfile structure type) por el momento nos quedaremos con aquella que el programa proporciona por default, a saber, Datos – Frecuencia regular (Dated – regular frequency) (ver Cuadro 1.2).
En cuanto a la especificación de los datos el programa proporciona diversas opciones desplegando la lista de la opción frecuencia (Frequency), entre ellas: datos de frecuencia anual (Annual), datos semestrales (Semi-annual), datos trimestrales (Quarterly), datos mensuales (Monthly), datos semanales (Weekly), datos diarios con semanas de 5 días (Daily – 5 day week), datos diarios con semanas de 7 días (Daily – 7 day week) y datos sin frecuencia definida (Integer date) (ver Cuadro 1.3).

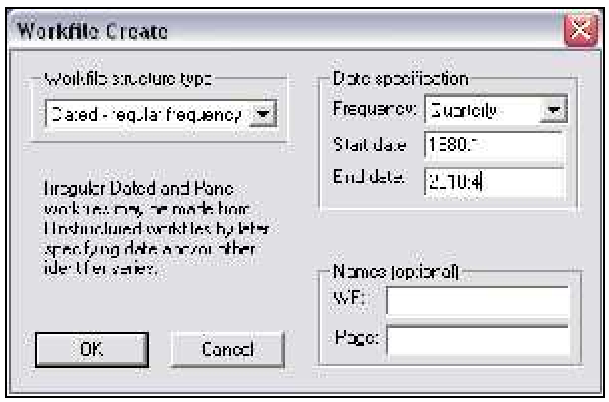
Para nuestro caso, como se trata de datos con frecuencia trimestral seleccionaremos la opción "Quarterly" y en fecha de inicio (Start date) escribiremos el primer año de nuestra serie seguido de dos puntos y el número 1, el cual indica que se inicia en el primer trimestre, si fuera a partir del segundo trimestre entonces se escribiría el número 2 y así sucesivamente hasta el número cuatro que simboliza el cuarto trimestre de cada año. En la fecha de finalización (End date) escribiremos el último año de nuestra serie seguido de dos puntos y el número 4, mismo que indica que se trata del cuarto trimestre, en caso de que no se cuente con la información hasta el cuarto trimestre únicamente se sustituye dicho número por el que corresponda al trimestre en cuestión. Así, en "Start date" escribiremos 1980:1 y en "End date" 2010:4/ Ok., (ver Cuadro 1.4) con lo cual habremos generado un nuevo archivo de trabajo y se desplegará una ventana como la que se muestra en el Cuadro 1.5.

Cuadro 1.2. Creación de un archivo de trabajo.

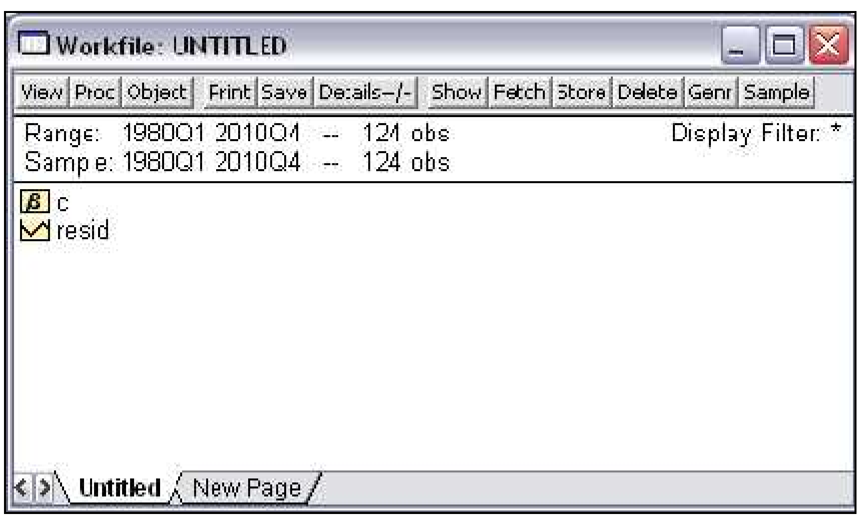
Cuadro 1.3. Frecuencia de los datos.



Cuadro 1.4. Especificación de los datos.



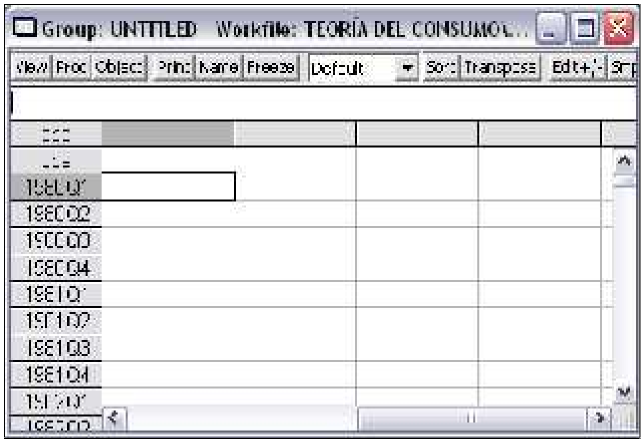
Cuadro 1.5. Archivo de trabajo.



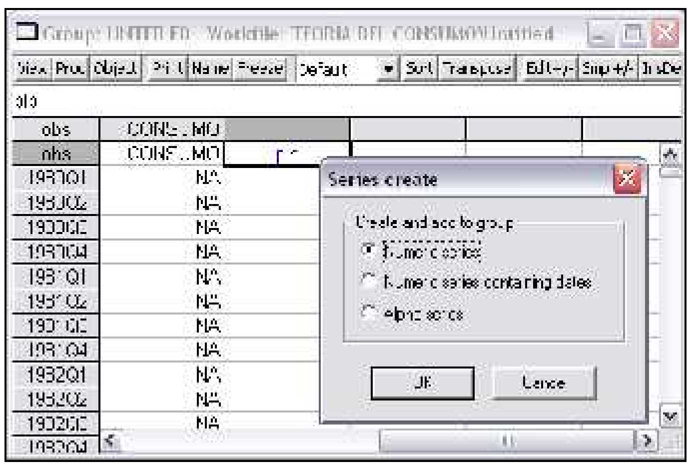
Como puede observarse en el Cuadro 5 el “Workfile” no tiene nombre o titulo ya que éste no ha sido guardado. Asimismo, esta ventana posee una barra de menú, seguida de una barra que indica el rango de los datos (Range) y la muestra de ese mismo rango (Sample). En este caso la muestra es igual al Rango, es decir, poseen el mismo periodo y por ende el mismo número de observaciones: 124 en total. Si deseáramos ampliar o disminuir el rango basta con dar un doble click sobre el mismo dentro de la ventana del Workfile; de forma similar, se deseáramos únicamente trabajar con una parte del rango, es decir, con una muestra de éste, entonces hay que dar un doble click sobre el Sample y en la ventana que despliega escribir el rango de la muestra que se va a utilizar. Posteriormente, el archivo de trabajo genera dos objetos por default, por un lado, un objeto coeficiente de vectores (β) y, por otro, un objeto serie titulado “residuos” (Resid). Para guardar el Workfile vamos a la barra de menú del programa, a File/ Save as/ y en la ventana que despliega seleccionaremos el lugar en donde deseamos guardar el archivo, se le pondrá un nombre y deberemos asegurarnos que el tipo de archivo sea con terminación “wf1” lo que indica que se trata de un archivo de trabajo (Workfile). En nuestro caso el archivo se llamará “Teoría del consumo”. En caso de que no se muestre la terminación “wf1” deberemos volver a la ventana del archivo de trabajo seleccionándola dando un click sobre la barra de titulo y volver a seguir los pasos antes descritos. Al realizar la operación de guardar, el programa nos desplegará una ventana cuestionando sobre la conservación de los datos en el archivo de trabajo, la opción “Single precision” creará archivos más pequeños en el disco pero guardará los datos con menos dígitos de precisión, por lo que se corre mayor riesgo en cuanto a perder información o dañarse el archivo, por su parte “Double precision” creará archivos más grandes en el disco pero guardará los datos con más dígitos de precisión (16 frente a 7).

Por otra parte, <u>existen varias formas para ingresar los datos</u>: 1) Desde menú; y 2) Desde comandos. En cuanto al menú, vamos a Quick/ Empty Group (Edit Series), con lo cual se desplegará una ventana tipo Excel en la cual podrán capturarse o pegarse los datos (ver Cuadro 1.6). Esta ventana es un objeto grupo, misma que posee una barra de menú y un área para capturar los datos. Para <u>colocar el nombre a las series</u> hay que desplegar completamente hacia arriba la barra desplazadora de la derecha, después colocáremos el cursor en la primera celda en blanco con el nombre de “Obs” y <u>escribiremos la palabra “consumo”</u> seguido de oprimir la tecla enter, el programa devolverá una ventana cuestionando sobre el tipo de objeto que se va a crear, si se trata de una serie numérica (Numeric series), series numéricas que contienen fechas (Numeric series containing dates) o series alfa (Alpha series), en este caso seleccionaremos la primera opción y por tanto la primera columna se llama “consumo”; en seguida realizamos el mismo procedimiento pero ahora nombráremos a la columna 2 como “PIB” (ver Cuadro 1.7). Automáticamente el programa generará las variables de forma <u>individual</u> como objetos <u>serie</u>, mismos que aparecerán en el Workfile, si nosotros deseamos guardar las series pero como <u>objeto grupo</u>, es decir, de forma conjunta, entonces en la ventana del objeto grupo en la barra de menú vamos a Name/ y en la ventana que despliega, “Object name”, escribimos el nombre del grupo (Name to identify object), en este caso nos quedáremos con el nombre que el programa proporciona por default “group01” mismo que aparecerá en el Workfile (ver Cuadro 1.8 y 1.9)

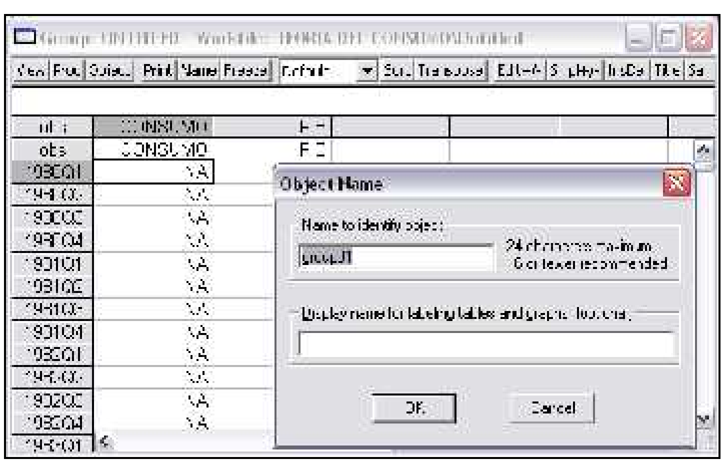
Cuadro 1.6. Ventana de datos.



Cuadro 1.7. Creación de series.



Cuadro 1.8. Guardar el objeto grupo.

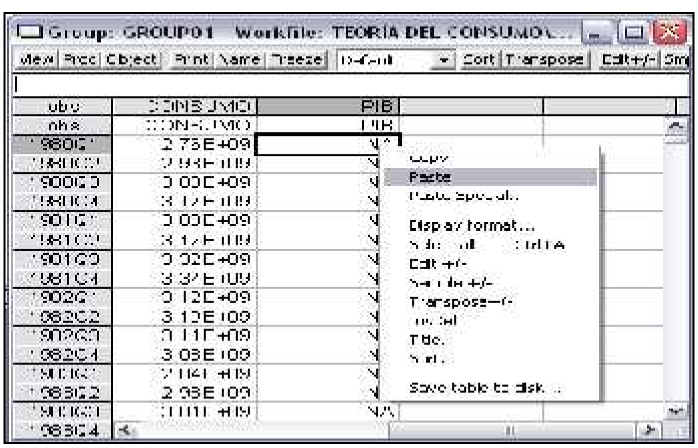


Cuadro 1.9. Estructura del Workfile.

En nuestro caso, para colocar los valores de las series Consumo y PIB los copiáremos del archivo de la base de datos en Excel y los pegáremos en la ventana del objeto grupo a partir de la fecha 1980Q1 dando un click derecho y seleccionando la opción "Paste" (ver Cuadro 1.10). Con ello se estará ya en condiciones de analizar el comportamiento de cada una de las series, la relación entre ellas y los resultados del análisis de regresión.

Por otra parte, para capturar las series desde la barra de comandos, en ésta escribiremos "data consumo PIB" con lo cual se desplegará una ventana como la que se muestra en el Cuadro 1.8, sin necesidad de escribir el nombre de las variables en cada una de las columnas como en la forma anterior. A continuación se copian y pegan los valores de las series siguiendo los pasos anteriormente descritos.

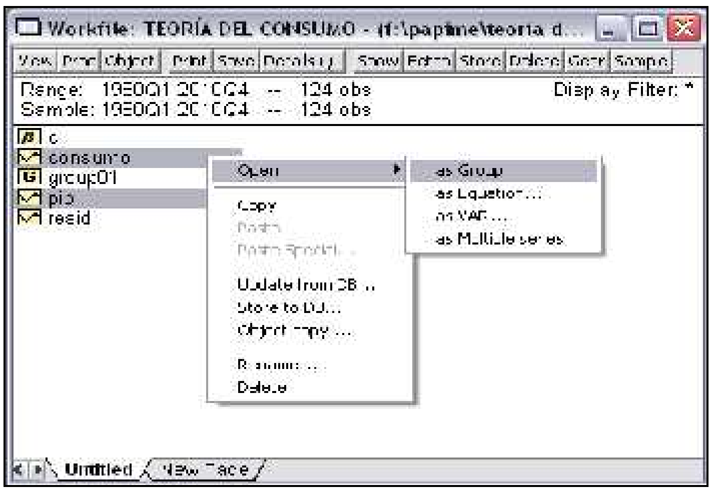
Cuadro 1.10. Ingresar datos.



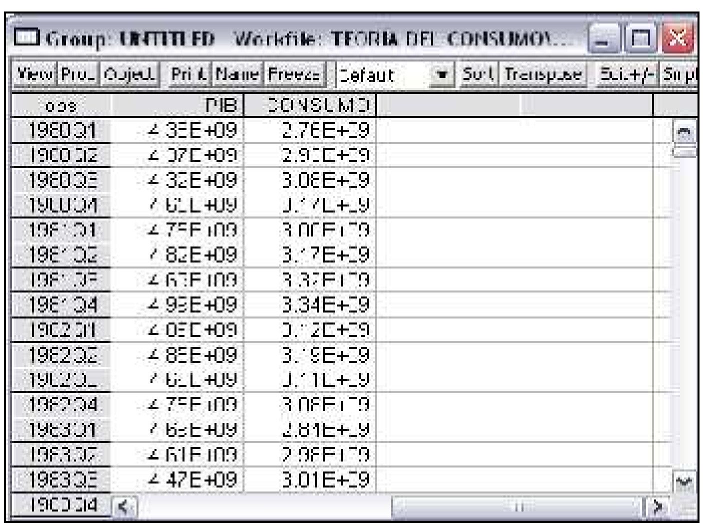
Una vez que ya están creadas las series (como se muestra en el Cuadro 1.9) podemos realizar el análisis gráfico y estadístico de las mismas. Para ello, desde la ventana del workfile seleccionamos primero la serie "PIB", por tratarse de la variable independiente, dando un click izquierdo sobre ella en el workfile y después seleccionamos la serie "Consumo", por tratarse de la variable dependiente, oprimiendo la tecla "control" y dando un click sobre la misma. A continuación, damos un click derecho dentro de la selección anterior, con lo cual se desplegará una lista de opciones, vamos a open/ as Group, (ver Cuadro 1.11), con lo que abriremos de forma conjunta las dos series (ver Cuadro 1.12). Lo mismo puede obtenerse dando un doble click sobre el objeto grupo (group01) que guardamos en el workfile.

-

Cuadro 1.11. Apertura de series.

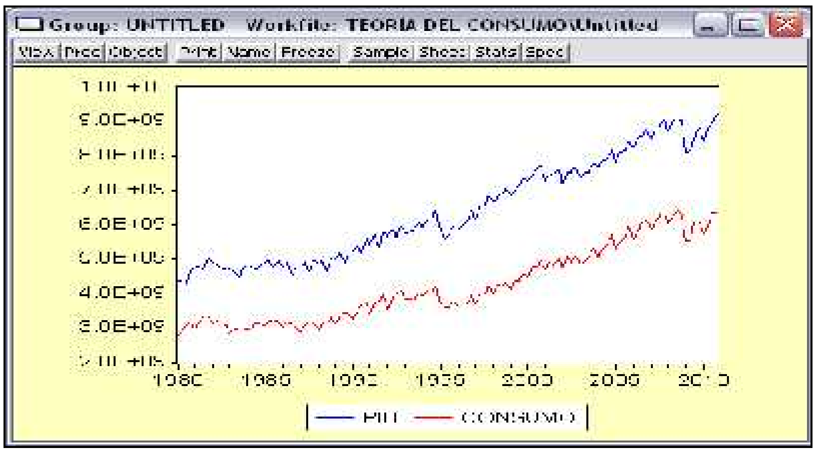


Cuadro 1.12. Ventana de datos.



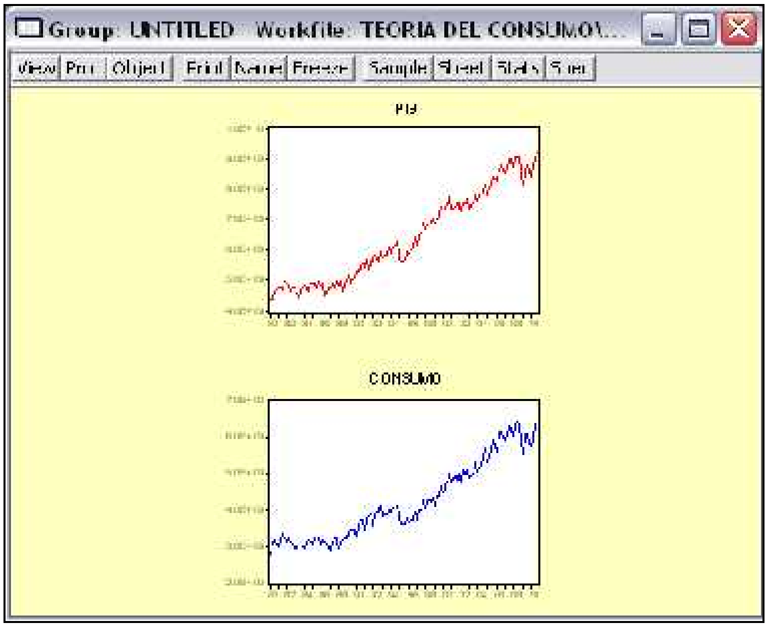
Ahora, sobre la ventana de datos (Cuadro 1.12) vamos a View/ Graph/ Line, con lo que se obtendrá una ventana con las gráficas conjuntas del PIB y el Consumo (ver Cuadro 1.13). No obstante, el programa proporciona una amplia variedad de tipos de gráficos: lineales (Line), de áreas (Area), de barras (Bar), etc., así como gráfica de forma individual, para ello en la ventana de los datos vamos a View/ Multiple Graphs/ Line, con lo cual se obtendrá una ventana como la que se muestra en la gráfica 1.1.

Gráfica 1.1. Gráfica lineal del PIB y el Consumo Privado de México, 1980-2010.
(Miles de pesos a precios constantes de 2003)



En caso de que deseáramos modificar el tipo de gráfico, el color, el grosor, etc., dentro de la gráfica que obtuvimos previamente damos un click derecho, lo que desplegará una lista de opciones, seleccionamos “options...”, lo que desplegará una ventana como la que se muestra en el Cuadro 13. Ahí podemos seleccionar el tipo de gráfico (Type), el color de fondo del gráfico, el tamaño, etc., (General), la leyenda de los datos (Legend), los atributos del gráfico, el color, el símbolo, etc., (Lines and symbols), etc. Asimismo, si deseamos incluir en el gráfico un Cuadro de texto, dentro del gráfico damos un click derecho y seleccionamos la opción “Add text”, lo que desplegará una ventana de dialogo en la cual deberemos escribir el texto deseado (Text for label), seleccionar la justificación del mismo (Justification) y la posición (Position). También, el software permite añadir dentro de la gráfica líneas para sobresaltar algunas observaciones o valores en particular, para ello dentro del gráfico damos un click derecho y seleccionamos la opción “Add shading” en la cual se podrá seleccionar la orientación de la línea (Orientation), vertical para una observación y horizontal para un valor sobre el eje de las ordenadas, al seleccionar una orientación deberemos escribir la posición de la misma (Position) y seleccionar el tipo de línea, el color de la misma, el grosor, etc.

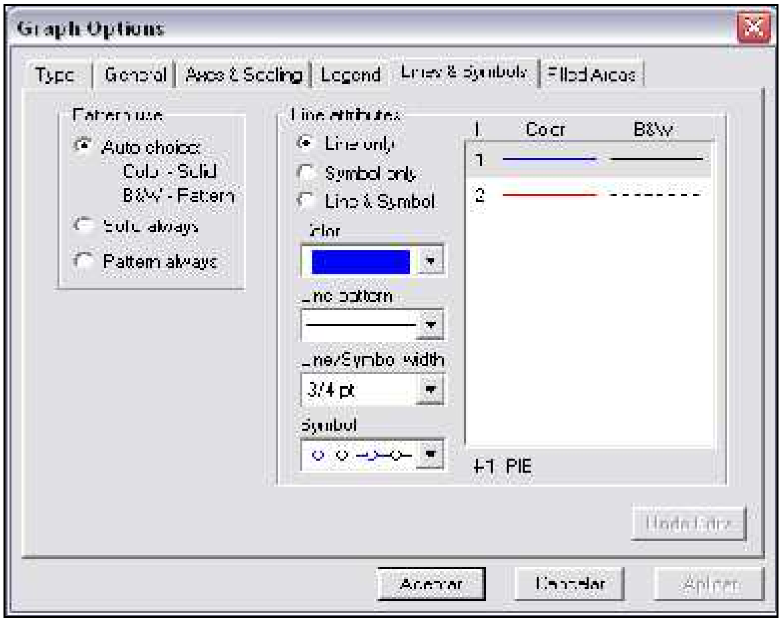
Gráfica 1.2. Gráficas individuales de las series.



En cuanto a las gráficas individuales (gráfica 1.2), en caso de querer modificar a ambas en cuanto al tipo de línea, color, grosor, añadir texto, líneas para resaltar observaciones o valores, etc., dentro del gráfico (área amarilla) damos un click derecho, lo que desplegará una lista de opciones, entre ellas: opciones para todas las gráficas (Options on all graphs), agregar o añadir líneas para todas las gráficas para resaltar observaciones o valores (Add shading to all graphs), y agregar texto (Add text). Si deseáramos únicamente modificaruna de las gráficas, entonces seleccionamos cualquiera de ellas y le damos un click derecho, con lo cual se desplegará una lista de opciones dividida en dos partes: 1) Para cambios en la selección; y 2) Para cambios en todas las gráficas, misma que ya describimos previamente. Para el caso de la primera, ésta proporciona opciones como: opciones para la selección, para la gráfica seleccionada (Options for selected), añadir líneas para resaltar observaciones o valores (Add shading to selected) y eliminar la selección (Remove selected), lo que eliminará la gráfica que hayamos seleccionado.

Así, como puede apreciarse en ambas gráficas tanto el ingreso (PIB) como el Consumo pueden ajustarse a un comportamiento lineal creciente de forma constante.

Cuadro 1.13. Opciones de gráfica.



Análisis Estadístico.

Para realizar el análisis estadístico de las series, una vez que éstas se han abierto conjuntamente (ver Cuadro 1.11 y 1.12) vamos a View/ Descriptive Stats/ Common Sample/ok con lo que se desplegará una ventana como la que se muestra en el Cuadro 1.14.

Cuadro 1.14. Análisis estadístico de las series.

	PIE	CONSUMO
Mean	6.40E+09	4.22E+09
Median	6.02E+09	3.95E+09
Maximum	9.24E+09	6.41E+09
Minimum	4.32E+09	2.76E+09
Std. Dev.	1.49E+09	1.09E+09
Skewness	0.326734	0.539739
Kurtosis	1.724807	2.713953
Jarque-Bera	10.60788	11.57094
Probability	0.004972	0.002922
Sum	7.93E+11	5.23E+11
Sum Sq. Dev.	2.72E+20	1.45E+20
Observations	124	124

La ventana anterior proporciona estadísticos como la media (Mean), la mediana (Median), el valor máximo (Maximum), el valor mínimo (Minimum), la desviación estándar (Std. Dev.), la asimetría (Skewness), la curtosis (Kurtosis), la estadística de normalidad Jarque-Bera (JB) con su probabilidad asociada (Jarque-Bera y Probability), la suma de todas las observaciones (Sum) y el número de observaciones (Observations) para cada una de las series.

De los datos anteriores se desprende que ninguna de las dos series, PIB y Consumo, son simétricas, pues la media no es igual a la mediana. Asimismo, éstas son platocúrticas ya que la curtosis es menor a 3 y ambas están cargadas a la derecha de la media ya que la asimetría es mayor que cero, por lo que las series no se distribuyen como una normal. Por otra parte, la estadística Jarque-Bera (JB), misma que se abordará con mayor detalle más adelante, permite probar la normalidad de las series, para ello se establece la siguiente prueba de hipótesis:

Ho: La serie se distribuye como una normal.

Ha: La serie no se distribuye como una normal.

La JB viene dada por:

$$JB = n\left[\frac{s^2}{6} + \frac{(k-3)^2}{24}\right]$$

Donde:

n = Número de observaciones.
s = Asimetría.
k = Kurtosis.

<u>En términos ideales la estadística</u> JB $\cong$ 6, por lo que, valores <u>iguales o mayores</u> a 6 en JB indican que la serie no se distribuye como una normal; en sentido contrario, valores en JB <u>menores a</u> 6 indican que la serie se distribuye como una normal.

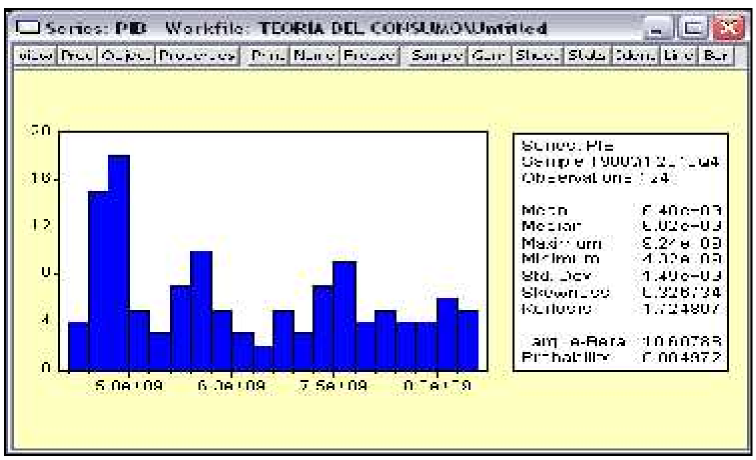
El programa proporciona una probabilidad asociada a la JB, misma que estableciendo α = 5%, indica la probabilidad mínima a la cual se rechaza la Ho. Así, si la probabilidad asociada a JB es igual o menor a 0.05 se rechaza la Ho y si es mayor a 0.05 <u>no s</u>e rechaza dicha hipótesis Ho.

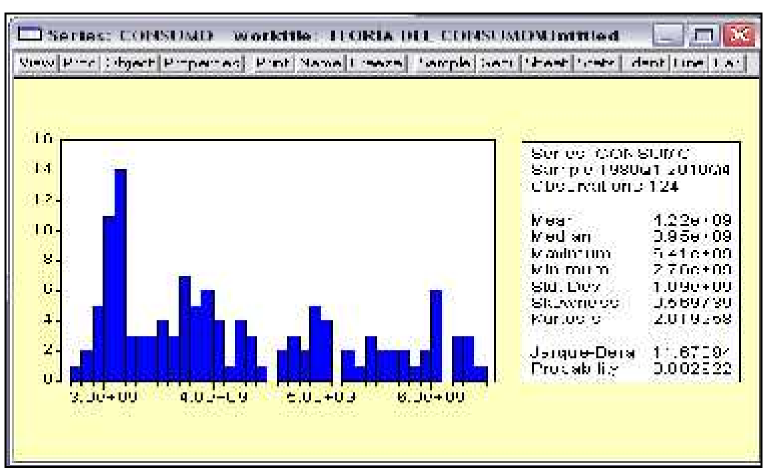
Por lo que, derivado del valor de la estadística JB y de su probabilidad asociada, las series PIB y Consumo no se distribuyen normalmente, es decir, se rechaza la Ho. La ausencia de normalidad en las <u>series nos alerta sobre el incumplimiento de algunos supuestos del método de estimación al relacionarlas.</u> Sin embargo, ello no implica que las series no puedan relacionarse mediante el análisis de regresión y correlación, <u>pues las pruebas sobre las violaciones a los supuestos del método</u>

de mínimos cuadrados se realizan sobre los residuales obtenidos a través de la regresión y no sobre las series originales. Así, el estudiante tendrá que poner especial atención en la correcta especificación del modelo en cuanto a la forma funcional, tamaño de la muestra, omisión de variables relevantes, inclusión de variables redundantes, etc.

El software también permite realizar el análisis estadístico de forma individual, generando además el histograma. Para ello, abrimos la serie dando doble click sobre ella en el workfile, en la ventana de la misma vamos a View/ Descriptive Statistics/ Histogram and stats, con lo cual se desplegará una ventana como la que se muestra en el Cuadro 1.15. Como se aprecia en dicho Cuadro, los resultados son los mismos que con el procedimiento anterior, sólo que ahora también nos proporciona un histograma para determinar visualmente la normalidad de las series. En este caso se corrobora el rechazo de la Ho planteada con anterioridad pues los histogramas no muestran una gráfica simétrica con forma de campana.

Cuadro 1.15. Histograma y estadísticas de las series.

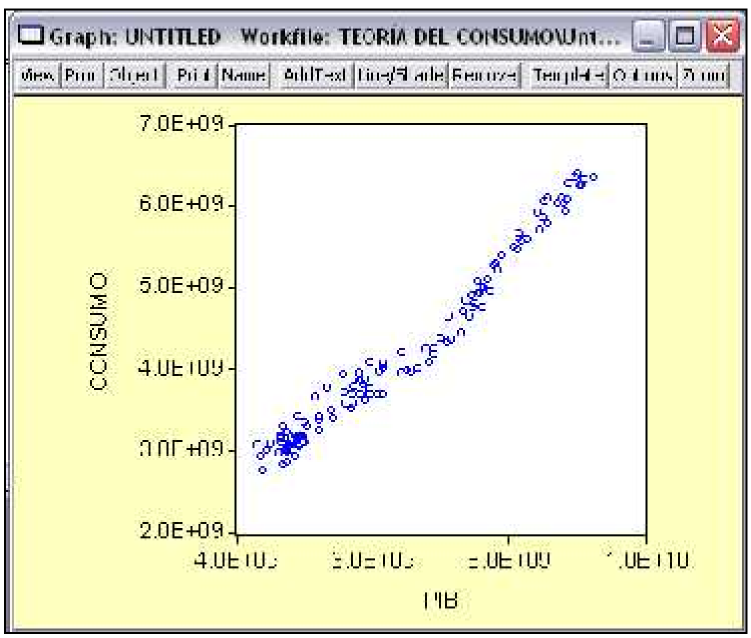




<u>Selección de la forma funcional con el diagrama de dispersión.</u>

Ahora bien, una vez realizado el análisis gráfico y estadístico, pasemos a determinar la forma funcional que corresponde a la relación entre el PIB (variable independiente) con el consumo (variable dependiente) a través del <u>diagrama de dispersión</u>. Para ello, existen distintas formas: 1) Desde menú; y 2) Desde comandos. Siguiendo la primera forma abrimos las series conjuntamente, seleccionando primero la independiente seguida de la dependiente, en ese orden para que la variable independiente se grafique en el eje de las abscisas (eje X) y la dependiente en el de las ordenadas (eje Y), en caso contrario el programa las graficará invertidas. En la ventana de las series vamos a View/ Graph/ Scatter/ Simple Scatter, con lo cual se generará una ventana como la que se muestra en la gráfica 1.3. También puede obtenerse el mismo resultado siguiendo los siguientes pasos: en la ventana de los datos vamos a View/ Multiple Graphs/ Scatter/ First series against all o en la barra del menú principal vamos a Quick/ Graph/ Scatter y en la ventana que devuelve escribimos el nombre de las series, primero la independiente seguido de la dependiente. Con la segunda opción, en la barra de comandos escribimos “scat pib consumo” seguido de la tecla enter.

Gráfica 1.3. Diagrama de dispersión.



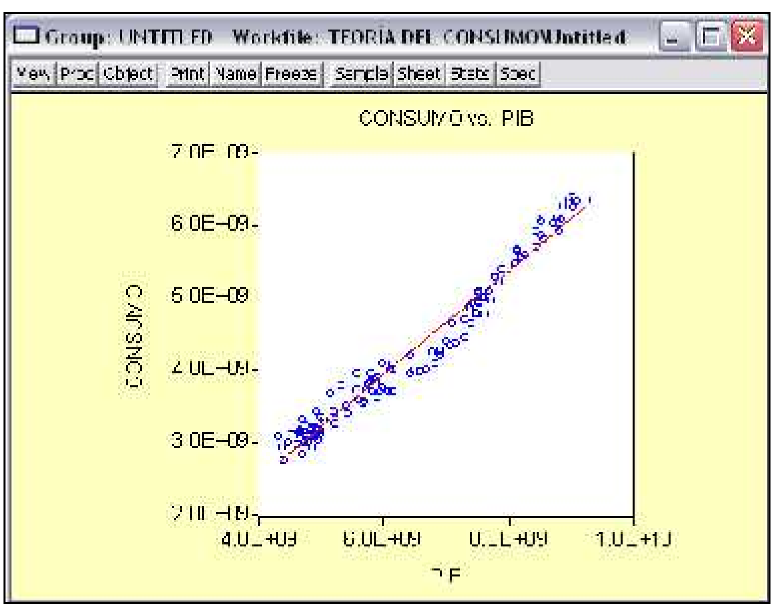
Como puede apreciarse la relación que siguen las variables se ajusta a una línea recta.

Para generar un diagrama de dispersión con regresión lineal, en la ventana de los datos vamos a View/ Graph/ Scatter/ Scatter with Regression y en la ventana que despliega “Global Fit Options” no modificamos nada, seleccionamos Ok., con lo cual se despliega una ventana como la que se muestra en la gráfica 1.4.

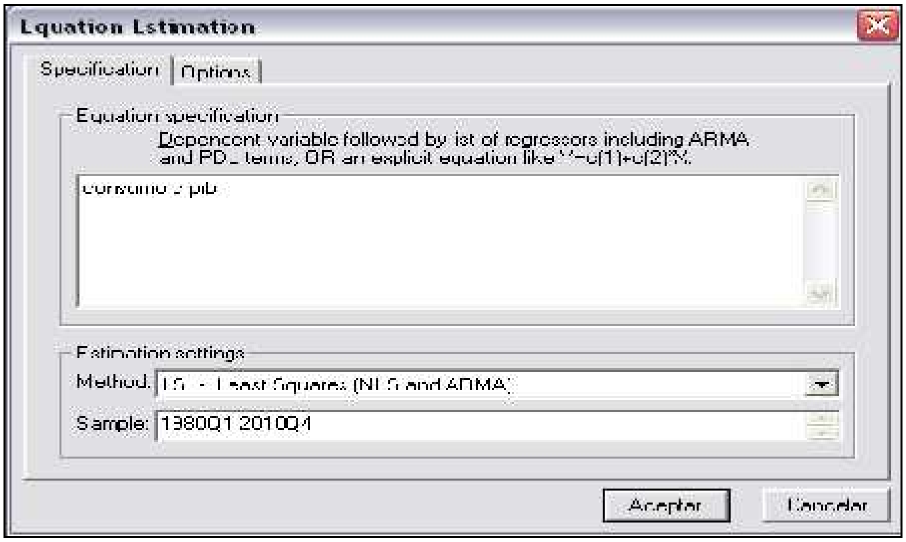
La ecuación de regresión.

Pasemos ahora a realizar la ecuación de regresión, para ello el programa proporciona dos opciones: 1) Por menú; y 2) Por comandos. Para el primer caso, en el menú principal del programa vamos a Quick/ Estimate Equation…, con lo que se desplegará una ventana como la que se muestra en el Cuadro 1.16. En la especificación de la ecuación (Equation specification) escribimos primero el nombre de la variable dependiente, en nuestro caso el Consumo, seguido de un espacio y la letra “c”, misma que denota la ordenada al origen, es decir, es la $\hat{a}$ de la ecuación de regresión que se planteó al inicio de este ejemplo, y después escribimos el nombre de la variable independiente, en nuestro caso el PIB. En configuración de la estimación (Estimation settings) el programa proporciona en método (Method) por default la opción de mínimos cuadrados (Least Squares-LS), sin embargo, proporciona una lista desplegable de opciones para llevar a cabo la estimación o diferentes modelos como son: Mínimos cuadrados en dos etapas (Two-Stage Least Squared-TSLS), Método de Momentos Generalizados (Generalized Method of Moments-GMM), Modelos Autoregresivos de Heteroscedasticidad Condicional (Autoregressive Conditional Heteroskedasticity-ARCH), Modelos de elección o respuesta binaria (Binary choice, como logit y probit), Modelos censurados o truncados (Censored or truncated data, Tobit), etc. También es posible modificar el tamaño de la muestra (Sample) que se utilizará para llevar a cabo la estimación. Apretamos el botón de aceptar y el programa devolverá una ventana como la que se muestra en el Cuadro 1.17.

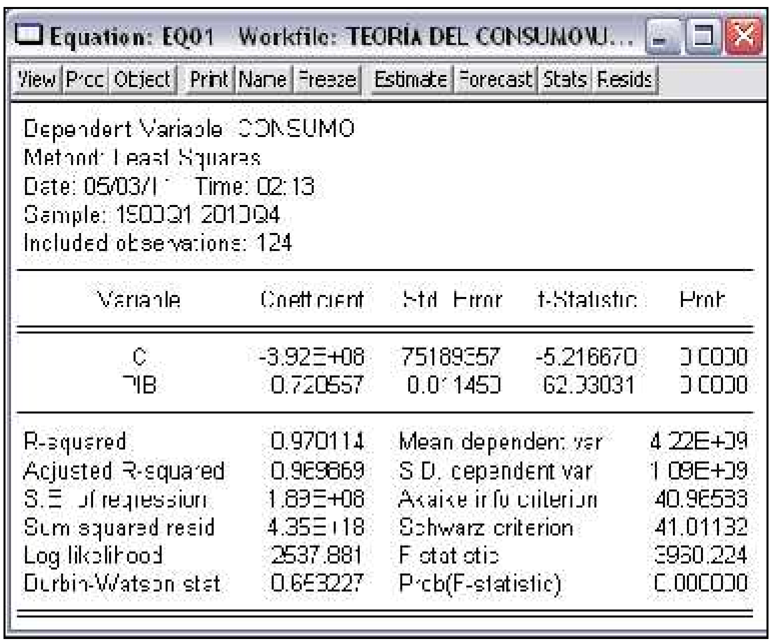
Gráfica 1.4. Diagrama de dispersión con regresión.



Cuadro1. 16. Estimación de la ecuación.



Cuadro 1. 17. Resultados de la ecuación de regresión.



Dependent Variable: CONSUMO
Method: Least Squares
Date: 05/03/11 Time: 02:13
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.92E+08	75189557	-5.216670	0.0000
PIB	0.720557	0.011450	62.93031	0.0000

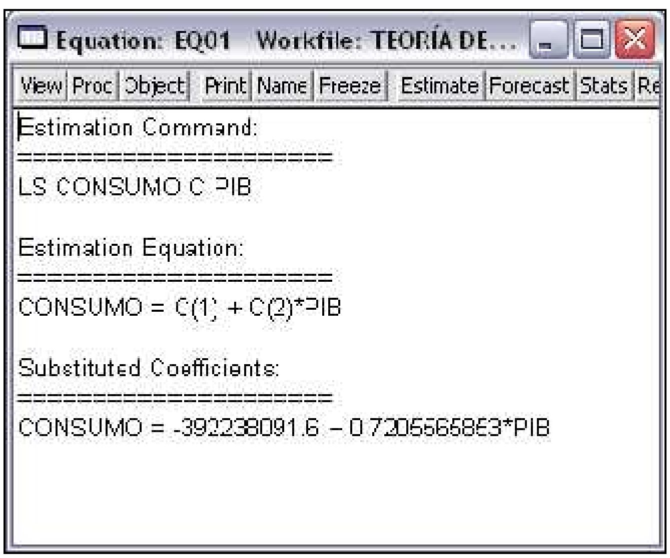
R-squared	0.970114	Mean dependent var	4.22E+09
Adjusted R-squared	0.969869	S.D. dependent var	1.09E+09
S.E. of regression	1.89E+08	Akaike info criterion	40.98533
Sum squared resid	4.35E+18	Schwarz criterion	41.01132
Log likelihood	2537.881	F-statistic	3960.224
Durbin-Watson stat	0.653227	Prob(F-statistic)	0.000000

Opción para crear la ecuación de regresión.

Asimismo, para generar la ecuación de regresión, desde el archivo de trabajo (workfile) seleccionamos primero la variable dependiente seguida de la variable independiente damos un click derecho dentro del área de la selección, lo cual desplegará una lista de opciones, seleccionamos Open/ As Equation, lo que devolverá una ventana como la que se muestra en el Cuadro 1.18.

La ventana mostrada en el Cuadro 1.17 contiene los resultados de la regresión. Así, en la parte superior se muestra el nombre de la variable dependiente (Dependent Variable), el método de estimación (Method), el día y la hora de creación (Date and Time), la muestra (Sample) y el número de observaciones (Included observations). Posteriormente en la parte central de la ventana de resultados se muestra el nombre de los estimadores (Variable), donde la “c” representa al estimador $\hat{a}$ y “PIB” al estimador $\hat{b}$, hacia la derecha se muestra el valor de los estimadores (Coefficient), seguido de su respectivo error estándar (Std. Error), el valor de su “t” estadística (t-Statistic) y la probabilidad asociada a la misma. En la parte inferior de la ventana de resultados se muestra el coeficiente de determinación (R^2), el coeficiente de determinación ajustado ($\overline{R}^2$), el error estándar de la regresión o estimación (S. E. of regression), la suma de los residuos al cuadrado (Sum squared resid), el logaritmo de verosimilitud (Log likelihood), la estadística Durbin-Watson (Durbin-Watson stat), la media de la variable dependiente (Mean dependent var), la desviación estándar de la variable dependiente (S. D. dependent var), el criterio de información de Akaike (Akaike info criterion), el criterio de Schwarz (Schwarz criterion), la “F” estadística (F-statistic) y su probabilidad asociada (Prob(F-statistic)).

Cuadro 1.18. Representación de la ecuación de regresión.



<u>Interpretación de los estimadores.</u>
Así, de acuerdo a los resultados de la estimación, la ecuación de regresión es:

$$\widehat{Consumo} = -392238091.6 + 0.7205 * PIB$$

Como se recordará, en teoría, $\hat{a}$ debe ser mayor que cero, pues representa el consumo mínimo (autónomo) que la sociedad debe tener, independientemente de su ingreso monetario. Por lo que al ser el coeficiente de dicho estimador negativo no se está cumpliendo con la teoría establecida, ello puede deberse a una mala especificación de la forma funcional, al tamaño de la muestra, a la omisión de variables explicativas, a la naturaleza de los datos, etc. No obstante, si se verifica que se cumple con todos los supuestos del método de estimación y dicho estimador continúa sin ser positivo se concluirá que la teoría así establecida no se cumple para el caso de México temporal y espacialmente, conclusión que puede ser totalmente aceptable, siempre y cuando se cumpla a cabalidad con los supuestos del método.

Ahora bien, en caso de que el valor del PIB fuera de cero, de acuerdo con la ecuación de regresión, el consumo autónomo, $\hat{a}$, será de – 392238091.6 miles de pesos constantes de
2003, puesto que:

$$Con\hat{s}umo = -392238091.6 + 0.7205 * (0) = -392238091.6 + 0 = -392238091.6$$

En cuanto al valor del estimador $\hat{b}$, observamos que éste si cumple con lo establecido en la teoría, pues su valor está contenido entre 0 y 1. Así, por cada mil pesos constantes de 2003 (unidades en las cuales está dado el PIB) que se incremente el ingreso (en este caso el PIB) el consumo se incrementará, dado su signo positivo, 0.7205 miles de pesos constantes de 2003 (unidades en las cuales está dado el consumo), o lo que es lo mismo 720.5 pesos constantes de 2003, y viceversa, es decir, si el ingreso disminuye en mil pesos constantes de 2003, el consumo disminuirá 720.5 pesos constantes de 2003. Por lo que, de acuerdo con la teoría diremos que por cada unidad monetaria adicional de ingreso en la sociedad, el consumo de ésta, en general, se incrementa en 0.7205 unidades monetarias y viceversa.

En este contexto, la Propensión Marginal a Consumir (PMgC) de la sociedad mexicana es de 0.7205 y la Propensión Marginal a Ahorrar (PMgS) es igual a:

$$PMgS_{México} = 1 – PMgC = 1 – 0.7205 = 0.2795$$

Lo que significa que por cada unidad monetaria adicional de ingreso obtenido, 0.7205 se destinan al consumo y 0.2795 a ahorrar.

Cabe aclarar que en este primer acercamiento (regresión simple) únicamente se pretende ilustrar una pequeña parte del método econométrico como herramienta de gran utilidad en el análisis económico, ya que en este momento no se han realizado, ni se realizarán, las pruebas de violación a los supuestos del método de estimación, y menos aún la corrección de ello. De ahí que hagamos la interpretación de los resultados y arribemos a conclusiones parcialmente aceptables y/o verdaderas, ya que los resultados hasta ahora obtenidos no pueden ser definitivas y generalizables, y mucho menos, utilizados para la toma de decisiones, puesto que desconocemos aún si se cumplen o no los supuestos detrás del método de estimación[12].

A continuación procedemos a realizar la prueba de significación estadística de los estimadores. Para ello establecemos la siguiente prueba de hipótesis:

Ho: $\beta = 0$, por tanto, el ingreso (PIB) no explica al consumo.
Ha: $\beta \neq 0$, por tanto, el ingreso (PIB) explica al consumo.

Como se recordara, en la práctica no se prueba la significancia estadística de la ordenada al origen, $\hat{a}$, ya que ésta no tiene asociada ninguna variable explicativa, es decir, es independiente y representa tan sólo un punto de partida. Sin embargo, la decisión de probar su significancia estadística también dependerá de la teoría que se esté probando. Así, por ejemplo, al ser $\hat{a}$ el consumo autónomo en la teoría keynesiana del consumo y al estar sometida ésta a una restricción, probar su significancia estadística es de suma importancia.

Para probar la significancia estadística de los estimadores, en este caso el de la variable explicativa, es decir, el PIB, se compara la "t" calculada (t_β) con la "t_α", también conocida como "t" teórica o de tablas. Por su parte, la "t" calculada se obtiene de la siguiente manera:

$$t = \frac{\hat{b} -}{s_{\hat{b}}} = \frac{0.720557 - 0}{0.011450} = 62.9303$$

El resultado obtenido de la "t_β" puede corroborarse en la ventana de los resultados de la ecuación de regresión (ver Cuadro 1.17). Dicho resultado se compara con la "t_α" con un nivel de significación estadística α=5% y grados de libertad igual a n-k=124-2=122, cuyo valor en tablas es igual a $t_\alpha = \pm 2.2694$.

Los criterios de decisión para rechazar o no la Ho son los siguientes:

1. Si t $\geq t_\alpha$, ambas en términos absolutos, entonces se rechaza la Ho.
2. Si t $< t_\alpha$, ambas en términos absolutos, entonces no se rechaza la Ho.

Así, al ser t = 62.9303 > t_α = 2.2694 se rechaza la Ho, por lo que la variable independiente explica a la dependiente, en nuestro caso, el ingreso (PIB) explica al consumo al 95% de probabilidad de que así sea.

[12] Como se verá más adelante, el incumplimiento a los supuestos del método de estimación genera estimadores sesgados, ineficientes, inconsistentes, etc., lo cual conlleva a que los resultados obtenidos no puedan ser aceptados y utilizados con fines de análisis de estructura, de predicción y de toma de decisiones.

El programa también proporciona la probabilidad asociada al estadístico "t" (p-value), la cual representa la probabilidad máxima a la cual se rechaza la Ho. Los criterios de decisión, con un nivel de significación al 5%, para rechazar o no la Ho con el p-value son:

1. Si la probabilidad asociada al estadístico "t" es menor o igual a 0.05 se rechaza la Ho.
2. Si la probabilidad asociada al estadístico "t" es mayor a 0.05 no se rechaza la Ho.

Por lo que en nuestro caso se corrobora que se rechaza la Ho ya que la probabilidad asociada a "t" es menor a 0.05.

En cuanto a la prueba de significancia estadística de $\hat{a}$, se plantea la siguiente prueba de hipótesis:

Ho: α = 0, por tanto, su valor no tiene o no es de relevancia económica.
Ha: α > 0, por tanto, su valor si es de relevancia económica.

Como la probabilidad asociada a t_a es menor a 0.05 se rechaza la Ho, lo que significa que el valor de $\hat{a}$ es estadísticamente significativo y por ende de relevancia económica. <u>Sin embargo, el valor que asume es negativo, lo cual no cumple con la teoría y la lógica económica.</u>
<u>¿ Qué hacemos en este caso? Respuesta: aplicar procedimientos tales como la ampliación del tamaño de la muestra, deflactar los datos de las variables, transformar los valores de las variables a logaritmos, etc ya que es muy importante corregir el signo de la varaiable en esta teoría económica</u>

La prueba de significación estadística <u>global</u> del modelo en regresión simple no se realiza, ya que únicamente existe <u>una</u> variable independiente. Por lo que, nadamás diremos que F fue calculada en el Cuadro 1.17, en regresión simple, <u>que no se usa</u> y que su valor es:

$$F = t^2$$
$$F = (62.93031)^2 = 3960.224$$

<u>*Bondad de ajuste.*</u>

Por lo que respecta a la prueba de bondad de ajuste, el coeficiente de determinación (R^2) es de 0.970101, que al multiplicarse por 100 es igual a 97.0101%. <u>Lo que indica que hay causalidad o dependencia de la siguiente magnitud:</u> 97.0101% de los cambios o variaciones en el consumo se deben o explican por cambios o variaciones en el ingreso, y el resto se explica <u>por variables</u> que <u>no fueron incluidas explícitamente</u> en el modelo.

El coeficiente de determinación ajustado ($\bar{R}^2$), el cual se ajusta mediante la introducción de grados de libertad, determina la variación que es explicada por la variable independiente, con respecto a la variable dependiente, cuando se introduce una variable independiente adicional al modelo. Éste viene dado por:

$$\bar{R}^2 = 1 - \left[(1 - R^2) * \frac{n-1}{n-k} \right]$$

Donde:

$\bar{R}^2$ = Coeficiente de determinación ajustado.
R^2 = Coeficiente de determinación.
n = Tamaño de la muestra.
k = Número de coeficientes estimados.

Por lo que, sustituyendo R^2 = 0.970101, n = 124 y k =2, tenemos:

$$\bar{R}^2 = 1 - \left[(1 - 0.970101) * \frac{124-1}{124-2} \right]$$
$$\bar{R}^2 = 1 - \left[(0.029899) * \frac{123}{122} \right]$$
$$\bar{R}^2 = 1 - \left[(0.029899) * 1.008196 \right]$$
$$\bar{R}^2 = 0.969869$$

En nuestro caso $\bar{R}^2$ es de 0.969869, lo que significa que el 96.9869% de las variaciones en el consumo se explican por variaciones en las variables independientes, en este caso el PIB.

En cuanto al coeficiente de correlación parcial (r) este es igual a:

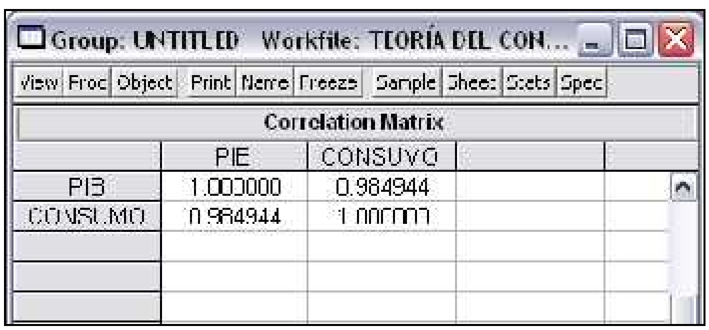
$$r = \pm\sqrt{R^2} = \pm\sqrt{0.970101} = +0.984944$$

Como se recordara “r” asume valores desde -1 hasta 1, pasando por el cero. El “r” *no mide causalidad o dependencia*, mide el grado de relación o asociación que existe entre dos variables, pudiendo tener éstas una relación inversa o negativa, cuando “r” tiene signo negativo, y directa o positiva, cuando “r” tiene signo positivo. En teoría se considerará un “r” aceptable cuando éste asume un valor mayor o igual a 0.5 en términos absolutos, trabajando con las series en sus valores originales (niveles). Así, para nuestro ejemplo, obtuvimos un r=0.984944, lo que significa que existe un alto grado de asociación entre las dos variables y además dicha relación es directa, dado el signo positivo, por lo que si una crece la otra también lo hará y viceversa. Cabe mencionar que al decidir el signo del resultado de la raíz cuadrada del R^2 nos quedamos con el positivo, en este caso, porque el signo obtenido en la ecuación de regresión del estimador de la variable independiente, el PIB, es positivo.

Correlación parcial.

Por su parte, EViews presenta “r” como una matriz de correlaciones parciales, para generarla existen dos formas: 1) Desde menú; y 2) Desde comandos. Para el caso del menú, una vez que se han abierto las series conjuntamente (ver Cuadro 1.12) vamos a View/ Correlations/ Common Sample, con lo cual se desplegará una ventana como la que se muestra en el Cuadro 1.19. Desde comandos, en la barra de comandos escribimos: cor pib consumo, seguido de oprimir la tecla enter, con lo que se desplegará una ventana igual a la del Cuadro 1.19. Como puede apreciarse la diagonal principal de la matriz de correlaciones es igual a 1, ya que se están relacionando las variables consigo mismas, aquí lo que interesa son los coeficientes de correlación parciales que están por arriba o por debajo de la diagonal principal de la matriz y que son exactamente iguales, ya que la matriz es simétrica. Así, el “r” entre el consumo y el PIB es exactamente igual al “r” entre el PIB y el consumo.

Cuadro 1.19. Matriz de correlaciones parciales.



	PIB	CONSUMO
PIB	1.000000	0.984944
CONSUMO	0.984944	1.000000

Observaciones: Por ser regresión simple vemos que ambos, los coeficientes de correlación general y parcial tienen el mismo valor porque sólo se usa una variable explicativa; en otras palabras, si fuera regresión multiple, ergo, que usaramos dos o más variables explicativas, sus valores difiririan, como se verá mása adelante.

Obtención de los valores estimados ($\hat{Y}$)

Finalmente, por lo que respecta a los valores estimados del consumo a través de la ecuación de regresión ($\hat{Y}$), estos se obtienen al sustituir los valores que asume la variable independiente, el PIB, en dicha ecuación. Así, por ejemplo, el valor que asume el PIB en el primer y segundo trimestre de 1980 es 4384924682.44 y 4372424964.63 miles de pesos constantes de 2003, respectivamente. Estos valores se sustituyen en la ecuación de regresión, obteniendo para el primer trimestre de 1980:

$$Con\hat{s}umo_{1980Q01} = -392238091.6 + 0.7205 * PIB$$
$$Con\hat{s}umo_{1980Q01} = -392238091.6 + 0.7205(4384924682.44)$$
$$Con\hat{s}umo_{1980Q01} = 2767348264.38$$

Para el segundo trimestre de 1980:

$$Con\hat{s}umo_{1980Q02} = -392238091.6 + 0.7205 * PIB$$
$$Con\hat{s}umo_{1980Q02} = -392238091.6 + 0.7205(4372424964.63)$$
$$Con\hat{s}umo_{1980Q02} = 2758341510.39$$

Errores o residuos.

Los errores o residuos se obtienen de restar al valor real de la variable dependiente, en este caso el consumo, el valor estimado de la misma mediante la ecuación de regresión, en este caso $Con\hat{s}umo$, es decir:

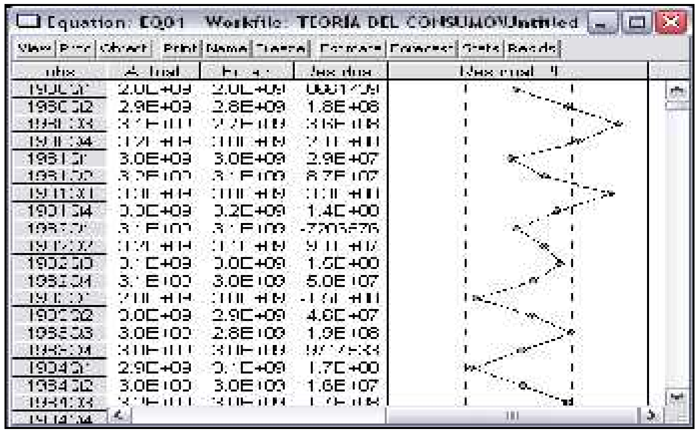
$$e_i = Y - \hat{Y} = Consumo - Con\hat{s}umo$$

Así, el error tanto para el primero como para el segundo trimestre de 1980 es:

$$e_{1980Q01} = Consumo - Con\hat{s}umo = 2758686525.75 - 2767348264.38 = -8661738.63$$
$$e_{1980Q02} = Consumo - Con\hat{s}umo = 2933851004.42 - 2758341510.39 = 175509494.03$$

En EViews los valores estimados (Fitted) de la variable dependiente, $Con\hat{s}umo$, y los errores (Residual) se obtienen estando en la ventana de la ecuación de regresión (ver Cuadro 1.19) en View/ Actual, Fitted, Residual/ Actual, Fitted, Residual Table, con lo que se desplegará una ventana como la que se muestra en el Cuadro 1.20. Es oportuno comentar que únicamente se muestra una parte de estos resultados por cuestiones de espacio.

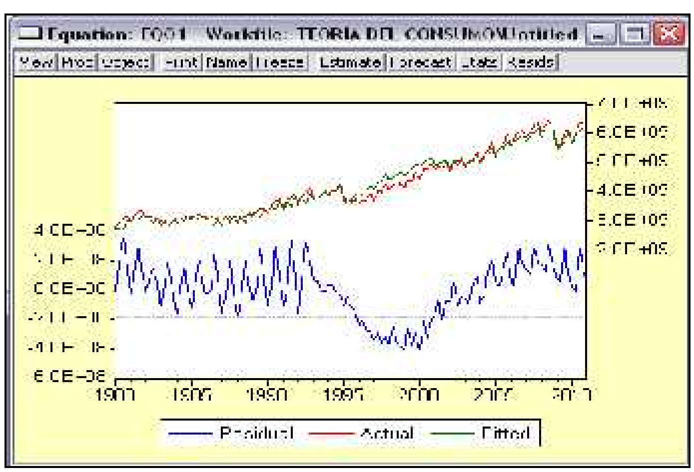
Cuadro 1.20. Valores reales y estimados de la variable dependiente y errores.



El Cuadro 1.20 presenta los valores reales u observados de la variable dependiente (Actual), los valores estimados de la misma mediante la ecuación de regresión (Fitted) y la diferencia entre ambos, es decir, los términos de error (Residual). Asimismo, grafica los términos de error con unas bandas de confianza (línea punteada) construidas a más/menos 2 desviaciones estándar.

Para obtener una <u>gráfica</u> de los valores observados y estimados de la variable dependiente y la diferencia entre ambos, en la ventana de la regresión vamos a View/ Actual, Fitted, Residual/ Actual, Fitted, Residual Graph, con lo que se desplegará una ventana como la contenida en la gráfica 1.5. La misma ventana puede ser generada oprimiendo el botón “Resids” en la ventana de la ecuación de regresión. En caso de que únicamente se desee graficar los términos de error, en la ventana de la ecuación vamos a View/ Actual, Fitted, Residual/ Residual Graph.

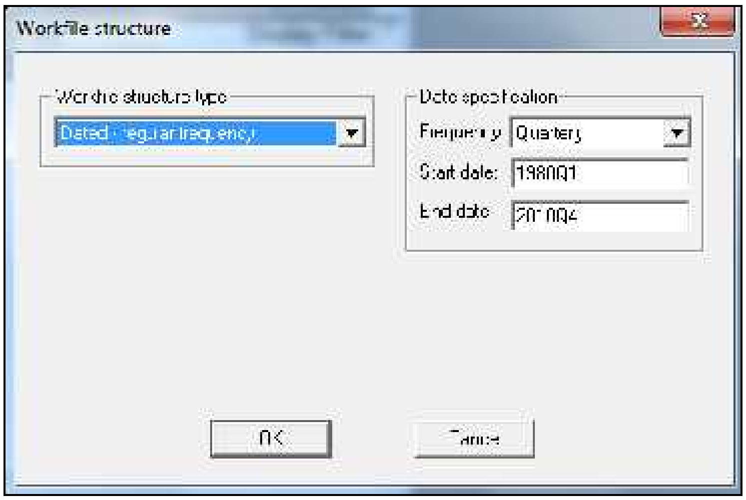
Gráfica 1.5. Representación de los valores reales y estimados de la variable dependiente y los errores.



IV.12.1.1.1.- Predicción.

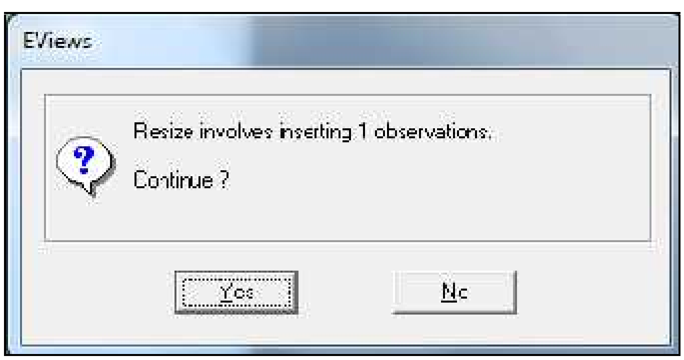
Ahora bien, como estamos usando la metodología de la econometría clásica, en este caso primero se obtiene el valor futuro de la variable independiente, por ejemplo supongamos que el valor del PIB para el primer trimestre de 2011 es de 9307060457.75. Así, en Eviews en la ventana del Workfile damos doble click sobre el Rango (Range), con lo cual se desplegará una ventana como la siguiente:

Cuadro 1.21. Estructura del Rango.



En esta ventana en “End date” modificamos el valor y ponemos “2011Q1”, con lo cual se incrementará el Rango original en una observación. El programa cuestiona sobre si verdaderamente desea incertar una observación adicional, mostrando la siguiente ventana:

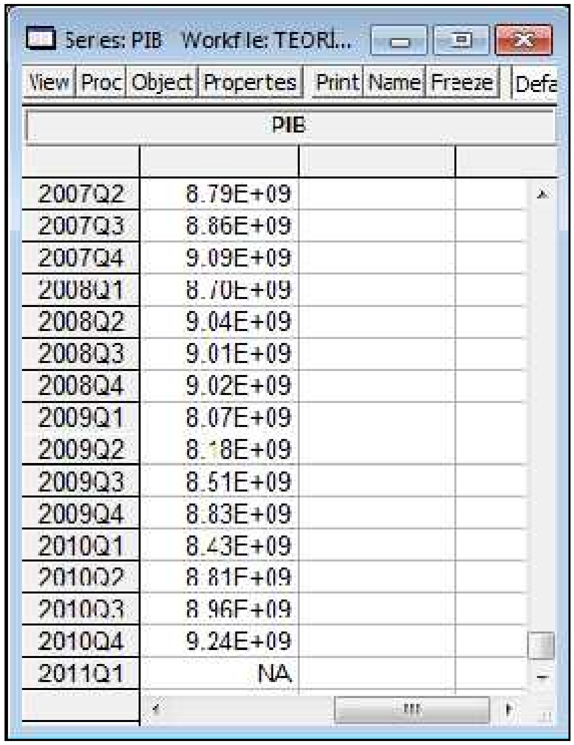
Cuadro 1.22. Incorporación de una observación.



En la ventana anterior seleccionamos la opción de “Yes”, con lo cual se se modificará el Rango. Cabe mencionar que con esta operación también se incrementa la muestra (Sample) en esa misma observación.

Una vez realizado el procedimiento anterior, en el Workfile seleccionamos y abrimos la serie “PIB” dando doble click sobre ella. Con ello se obtienen los valores de la serie comentada, nos desplazamos sobre dicha ventana hacia a bajo y observamos que ahora contiene una celda vacia, misma que representa el incremento que realizamos tanto al Rango como a la muestra (ver Cuadro 1.23).

Cuadro 1.23. PIB ampliado en una observación.



2007Q2	8.79E+09
2007Q3	8.86E+09
2007Q4	9.09E+09
2008Q1	8.70E+09
2008Q2	9.04E+09
2008Q3	9.01E+09
2008Q4	9.02E+09
2009Q1	8.07E+09
2009Q2	8.18E+09
2009Q3	8.51E+09
2009Q4	8.83E+09
2010Q1	8.43E+09
2010Q2	8.81E+09
2010Q3	8.96E+09
2010Q4	9.24E+09
2011Q1	NA

En el menú de la ventana anterior seleccionamos la opción “Edit +/-“ para incluir el valor del PIB en el primer trimestre de 2011 (2011Q1). Procedemos a escribir el valor de 9307060457.75 en la celda que aparece con la leyenda “NA”. Ya capturado el valor anterior el Cuadro de los valores de la serie PIB muestra el nuevo valor incluido para dicho trimestre de 2011 (ver Cuadro 1.24).

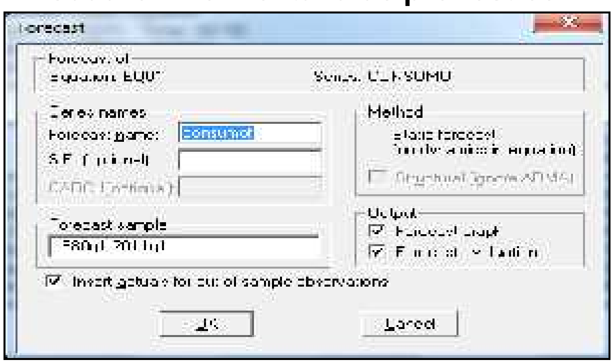
Realizada la operación anterior, el siguiente paso consiste en regresar a la regresión original (ver Cuadro 1.17), en la ventana de la regresión vamos al menú y seleccionamos la opción “Forecast” con lo que se desplegará una ventana como la contenida en el Cuadro 1.25. Observe el lector que el nombre de la serie pronostica es “consumof”, donde la “f” denota la palabra “Forecast” o predicción. El periodo de predicción (Forecast sample) corresponde desde el primer trimestre de 1980 hasta el primer trimestre de 2011. **Cabe señalar que de 1980 a 2010 los valores son estimaciones en el periodo histórico, mientras que para el primer trimestre de 2011 ya es un valor proyectado.** Seleccionamos la opción “Ok”, con lo que se desplegará una ventana como la contenida en el Cuadro 1.26.

Cuadro 1.24. Nueva composición de la serie PIB.

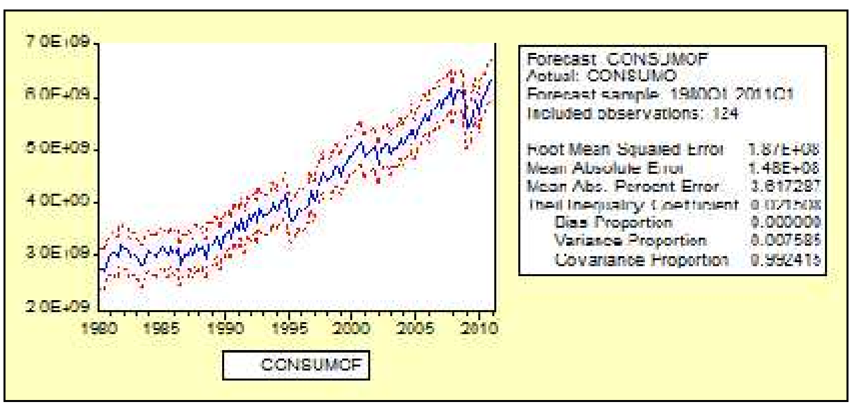
9307060457.75

		PIB
2006Q2	2006Q2	8.55E+09
2006Q3	2006Q3	8.56E+09
2006Q4	2006Q4	8.76E+09
2007Q1	2007Q1	8.60E+09
2007Q2	2007Q2	8.79E+09
2007Q3	2007Q3	8.86E+09
2007Q4	2007Q4	9.09E+09
2008Q1	2008Q1	8.70E+09
2008Q2	2008Q2	9.04E+09
2008Q3	2008Q3	9.01E+09
2008Q4	2008Q4	9.02E+09
2009Q1	2009Q1	8.07E+09
2009Q2	2009Q2	8.18E+09
2009Q3	2009Q3	8.51E+09
2009Q4	2009Q4	8.83E+09
2010Q1	2010Q1	8.43E+09
2010Q2	2010Q2	8.81E+09
2010Q3	2010Q3	8.96E+09
2010Q4	2010Q4	9.24E+09
2011Q1	2011Q1	9.31E+09

Cuadro 1.25. Ventana de pronóstico.



Cuadro y Gráfica 1.26. Bondad estadística de la proyección.



En el Cuadro 1.26 se muestra el estadístico de Theil entre otros indicadores estadísticos. Éste en particular nos permite determinar la bondad estadística de la proyección: Cuando Theil tiende a cero se dice que la proyección es confiable, lo cual se corrobora al observar en la gráfica que los valores estimados y el proyectado se encuentran dentro de las bandas de confianza.

La serie "Consumof" aparecerá contenida en la ventana del Workfile. Enseguida procedemos a compararla con la serie original del Consumo. Para ello en el Workfile seleccionaremos tanto la serie "Consumo" como "Consumof", las abrimos como grupo y aparece la ventana comparativa entre ellas (ver Cuadro 1.27).

Observese que la serie original del "Consumo" en la observación 2011Q1 aparece la leyenda "NA" dedibo a que esa observación no existe en la base de datos original. Mientras que en la serie "Consumof" la observación "2011Q1" aparece ya registrada, puesto que fue obtenida mediante la proyección anteriormente descrita.

Cuadro 1.27. Comparación de sus valores.

obs	CONSUMO	CONSUMOF
2007Q1	5.87E+09	5.73E+09
2007Q2	6.05E+09	5.94E+09
2007Q3	6.28E+09	5.99E+09
2007Q4	6.34E+09	6.16E+09
2008Q1	6.04E+09	5.88E+09
2008Q2	6.25E+09	6.12E+09
2008Q3	6.41E+09	6.10E+09
2008Q4	6.26E+09	6.11E+09
2009Q1	5.49E+09	5.42E+09
2009Q2	5.56E+09	5.50E+09
2009Q3	6.07E+09	5.74E+09
2009Q4	6.00E+09	5.97E+09
2010Q1	5.71E+09	5.69E+09
2010Q2	5.94E+09	5.96E+09
2010Q3	6.35E+09	6.07E+09
2010Q4	6.36E+09	6.27E+09
2011Q1	NA	6.31E+09

Con estas operaciones hemos realizado un ejercicio completo de regresión y correlación simple con Eviews, con el que vimos y por consiguiente nos familiarizamos con las características estadísticas de las variables que integran un modelo econométrico. Por consiguiente, enseguida mostraremos otros dos ejemplos para que el lector afiance su conocimiento sobre los aspectos básicos sobre cómo se formula una teoría económica, cómo se expresa matemáticamente y cómo se verifica estadísticamente.

IV.12.1.2.- Ejemplo 2: Hipótesis del crecimiento económico impulsado por las exportaciones.

Planteamiento de la teoría económica

Ante la crisis económica mundial de los años ochenta del siglo XX, la mayoría de los países adoptaron una nueva estrategia de crecimiento y desarrollo. Este nuevo paradigma orienta la economía hacia el exterior y promueve la liberalización comercial como herramienta fundamental para potenciar el crecimiento, considerando al sector exportador como la fuente dinámica e impulsora del mismo.

En este contexto, tomando como base la Ley de Thirlwall, en la cual explica que la demanda de exportaciones es fundamental para explicar las diferencias de crecimiento entre países, se formulará una hipótesis en la que las exportaciones afectan positivamente al crecimiento de la economía en los países que mantienen relaciones de comercio exterior. Asimismo, Michaely y Balassa(bibliografía), suponen que la liberalización comercial es un instrumento eficaz para promover el crecimiento de los países. Estos autores obtuvieron coeficientes de correlación significativos y positivos entre las tasas de crecimiento de las exportaciones y las tasas de crecimiento económico de diversos países (Cuadros, 2000: 38).

En teoría, al haber un aumento en las exportaciones puede ocurrir lo siguiente:

1. Darse un incremento en la producción real, ya que en otros países están demandando bienes y servicios, con esto se incentiva la demanda interna.
2. Promover la especialización de la producción para su exportación, esto será un factor para que aumente la productividad y se pueda dar la competitividad en el mercado internacional.
3. Con un aumento de las exportaciones facilita la importación de insumos para satisfacer la demanda interna.
4. Al crearse una política comercial orientada hacia el exterior se permite el acceso a las tecnologías avanzadas.

Por su parte, Helpman y Krugman postulan que las exportaciones podrían aumentar por las economías de escala debido al incremento de la productividad, el aumento de las nuevas exportaciones pueden permitir reducciones de costos, lo cual puede resultar en mayores ganancias al aumentar la productividad (Giles, 2000:264) Bibliografía. Con estos puntos se da soporte para apoyar la hipótesis de que las exportaciones pueden afectar de manera positiva el crecimiento económico de un país.

Así, en este modelo de regresión simple se tomará el período comprendido entre el primer trimestre de 1980 al cuarto trimestre de 2010, con esto se confirmará o rechazará la hipótesis del crecimiento económico impulsado por las exportaciones para el caso de México.

El modelo matemático se plantea de la siguiente manera:

$$Producto\ Interno\ Bruto = f(Exportaciones)$$

Esto quiere decir que el Producto Interno Bruto está en función de las exportaciones.

En tanto que el modelo econométrico sería de la siguiente manera:

$$PIB = \hat{a} + \hat{b}X + \hat{e}$$

La relación teórica entre el Producto Interno Bruto (PIB) y las exportaciones (X) es directa, es decir, si las exportaciones crecen se espera que el Producto Interno Bruto también lo haga, y viceversa, si las exportaciones disminuyen el Producto Interno Bruto también lo hará.

De acuerdo con la teoría $\hat{a}$ representa Producto Interno Bruto independiente, es decir, el que no está relacionado con las exportaciones, este mismo debe ser mayor a cero, ya que existen otros determinantes que impactan al Producto Interno Bruto como lo es el consumo privado, la inversión, etc., por tanto:

$$\hat{a} > 0$$

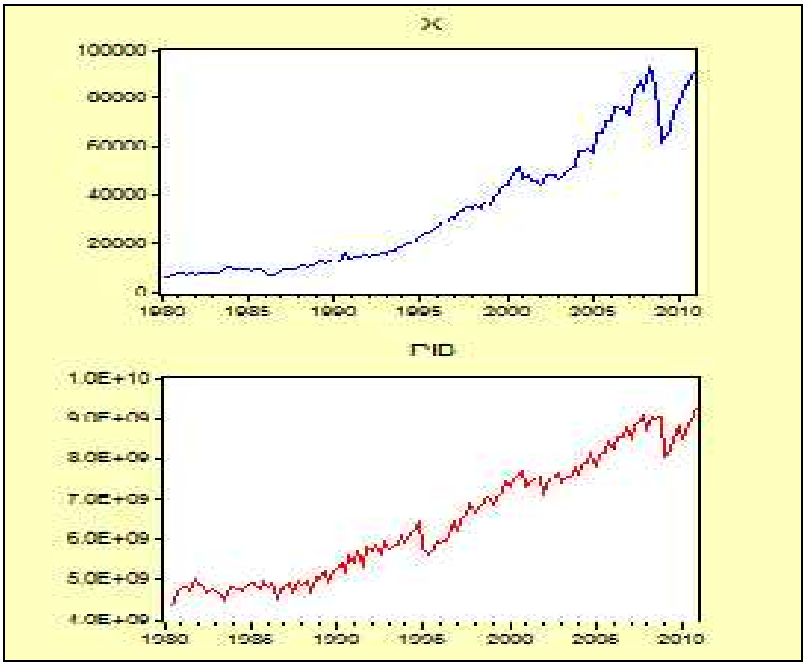
En cuanto a $\hat{b}$, éste es el multiplicador del Producto Interno Bruto (PIB) ante los cambios que pueden suscitarse dentro de las Exportaciones, en este caso se espera que éste sea positivo de acuerdo con la teoría.

$$\hat{b} \geq 0$$

A continuación se obtendrán las gráficas lineales del PIB y de las exportaciones, así como sus gráficas individuales (ver gráfica 2.1).

Ejemplo 2: Crecimiento impulsado por las exportaciones.

Gráfica 2.1. Gráficas individuales de las series.



En las gráficas individuales podemos observar que tanto el PIB como las exportaciones tienen pendientes crecientes. En el caso de las exportaciones se da un rápido crecimiento durante 1990 hasta el año 2000, ello debido a que durante la década de 1990 Estados Unidos tuvo un crecimiento promedio de 3.3%, por lo que se incrementó la demanda interna y se necesitaba abastecer estos mercados importando bienes y servicios. Esto fue ventajoso para México debido a la cercanía y las relaciones comerciales que se estaban formando, todo esto culminando con la firma del Tratado de Libre Comercio en 1994. Durante el período de 2000 a 2010 las exportaciones crecen pero no tan aceleradamente debido a que Estados Unidos sufrió dos crisis, en 2001-2002 y en 2008-2009, esta última crisis se originó en el sector financiero, afectando a la economía real reduciendo el consumo en todo el mundo y contrayéndose las exportaciones.

En cuanto al PIB se ven los efectos negativos dentro de 1995 debido al error de diciembre con la devaluación del peso causando una salida masiva de capitales del país y en 2008 debido a la crisis originada en Estados Unidos por la burbuja hipotecaria, que afectó de manera directa y en mayor proporción que otros países latinoamericanos debido a la fuerte dependencia de exportaciones e importaciones con aquel país.

Análisis estadístico inicial.

Por otra parte, tanto las curvas del PIB como las exportaciones no se distribuyen simétricamente, pues la media no es igual a la mediana. En cuanto al grado de picudez de la curva, son platocúrticas ya que la curtosis en el caso del PIB es $1.72<3$ y en las exportaciones $2.26<3$. Las series no son simétricas, siguen una asimetría positiva o cargada a la derecha ya que el coeficiente de asimetría es mayor que cero, por lo que las series no se distribuyen como una normal (ver Cuadro 2.1).

Con el estadístico Jarque-Bera (JB) se analizará la normalidad de las series, planteando la siguiente prueba de hipótesis:

Ho: La serie se distribuye como una normal.
Ha: La serie no se distribuye como una normal.

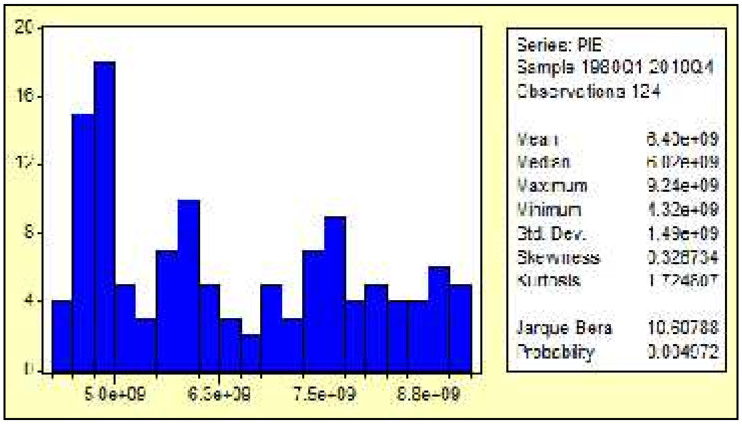
En términos ideales se estableció con anterioridad que JB $\cong 6$, en este ejemplo el PIB es de $10.60>6$ y en las exportaciones $14.02>6$, por lo que esto indica que las series no se distribuyen como una normal (ver Cuadro 2.1).

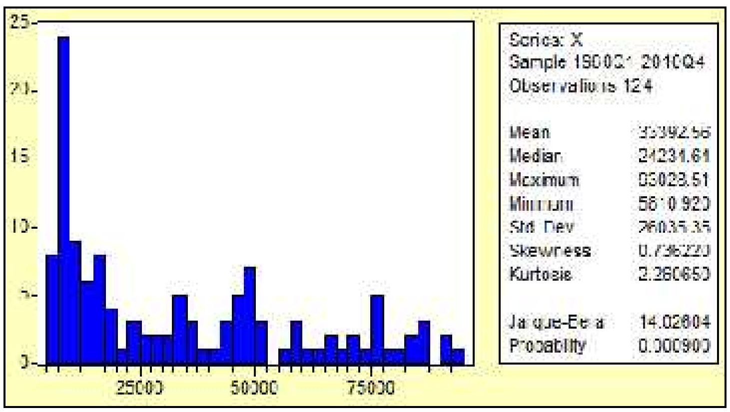
El programa proporciona una probabilidad asociada a dicha estadística que estableciendo $\alpha=5\%$, indica la probabilidad mínima a la cual se rechaza la Ho. Así, si la probabilidad asociada a JB es mayor a 0.05 no se rechaza la hipótesis nula y si es menor o igual a 0.05 se rechaza dicha hipótesis. En este ejemplo la probabilidad del PIB es $0.0049<0.05$ y en las exportaciones $0.0009<0.05$, lo cual significa que se rechaza la hipótesis nula, con lo cual se confirma que las series no se distribuyen como una normal (ver Cuadro 2.1).

Se observará de manera gráfica la normalidad de las series, ello mediante la información obtenida de los histogramas del PIB y las exportaciones. En este caso se rechaza la hipótesis nula planteada con anterioridad, ya que los histogramas no muestran una gráfica simétrica con forma de campana (ver Cuadro 1.29).

Ya realizado el análisis gráfico y estadístico para ver si hay normalidad en las variables, pasemos a determinar la forma funcional que corresponde a la relación entre las exportaciones (variable independiente) con el PIB (variable dependiente) a través del diagrama de dispersión. Como puede apreciarse en el Cuadro 2.1 la relación que siguen las variables se ajusta a una línea recta.

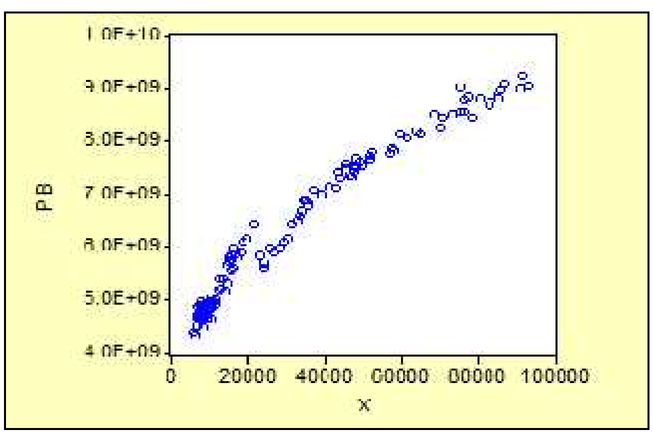
Cuadro 2.1. Histograma y estadísticas de las series.



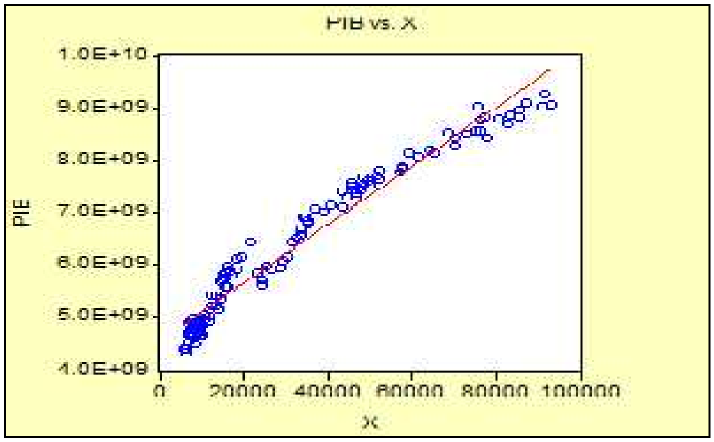


<u>Selección de la forma funcional:</u>

Gráfica 2.2. Diagrama de dispersión.



Gráfica 2.3. Diagrama de dispersión con regresión.



La recta de regresión (gráfica 2.2) muestra la forma en que el promedio del PIB aumenta conforme lo hacen las exportaciones, en este caso su relación parece ajustarse a una línea recta.

Creación de la ecuación de regresión

A continuación procedemos a generar la ecuación de regresión con EViews 5

Cuadro 2.2. Resultados de la ecuación de regresión.

Dependent Variable: PIB
Method: Least Squares
Date: 05/29/11 Time: 22:33
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.53E+09	44959134	100.8020	0.0000
X	55871.51	1063.415	52.53972	0.0000

R-squared	0.957674	Mean dependent var	6.40E+09
Adjusted R-squared	0.957328	S.D. dependent var	1.49E+09
S.E. of regression	3.07E+08	Akaike info criterion	41.93896
Sum squared resid	1.15E+19	Schwarz criterion	41.98445
Log likelihood	-2593.216	F-statistic	2760.422
Durbin-Watson stat	0.426927	Prob(F-statistic)	0.000000

De acuerdo a los resultados de la estimación en el Cuadro 2.2, la ecuación de regresión es:

*PIB = 4531969112 + 55871.51321*X*

Como se recordará, en caso de que el valor las de exportaciones fuera de cero, de acuerdo con la ecuación de regresión vemos que el PIB independiente del comportamiento de las exportaciones, ($\hat{a}$), será de 4531969112 miles de pesos constantes de 2003 puesto que:

$$PIB = 4531969112 + 55871.51321*(0)= 4531969112+0= 4531969112$$

<u>Verificación de la teoría económica.</u>

En cuanto al multiplicador del PIB ($\hat{b}$) tenemos que al incrementarse las exportaciones en un millón de dólares, el PIB crecerá en 55,871.51 miles de pesos constantes de 2003, sucederá lo contrario si disminuyen en un millón de dólares las exportaciones, en este caso el PIB se reducirá en 55,871.51 miles de pesos constantes de 2003. <u>Cabe resaltar la relación directa obtenida entre las X y el PIB, tal y como se esperaba en la teoría establecida con anterioridad.</u>

A continuación se procederá a realizar la prueba de significación estadística de los parámetros poblacionales. Para ello establecemos la siguiente prueba de hipótesis:

Ho: β = 0, por tanto, las exportaciones (X) no explican al Producto Interno Bruto (PIB).
Ha: β ≠ 0, por tanto, las exportaciones (X) explican al Producto Interno Bruto (PIB).

Para probar la significación estadística de los estimadores, se compara la "t" calculada del estimador (tc) con la "t_α". Esto de la siguiente manera:

Los criterios para rechazar o no la Ho son los siguientes:

1. Si $t_C \geq t\alpha$, ambas en términos absolutos, entonces se rechaza la Ho.
2. Si $t_C < t\alpha$, ambas en términos absolutos, entonces no se rechaza la Ho.

Así, al ser t = 52.53 > t_α = 2.2694 se rechaza la Ho, esto quiere decir que la variable independiente explica a la dependiente, es decir, las exportaciones explican al PIB al 95% de probabilidad de que así sea.

Analizando la <u>probabilidad</u> asociada al estadístico "t" (p-value), la cual representa la probabilidad mínima a la cual se rechaza la Ho. Los criterios, con un nivel de significación al 5% son:

1. Si la probabilidad asociada al estadístico "t" es menor o igual a 0.05 se rechaza la Ho.

2. Si la probabilidad asociada al estadístico "t" es mayor a 0.05 no se rechaza la Ho.

Tomando las probabilidades asociadas al estadístico "t" que proporciona Eviews 5, en el Cuadro 2.2 se obtuvo que en la constante fue de 0.0000<0.05 y en las exportaciones fue de 0.000<0.05, como en ambos casos es menor a 0.05 se rechaza la hipótesis nula.

Bondad de ajuste.

Para la prueba de bondad de ajuste tenemos que R^2=0.957674= 95.7674%, lo que significa que el 95.7674% de los cambios o variaciones en el Producto Interno Bruto se explican por cambios en las exportaciones. En cuanto al coeficiente de determinación ajustado $\overline{R}^2$= 0.957328= 95.7328% de los cambios en el Producto Interno Bruto se explican por los cambios en las exportaciones y una variable adicional.

El coeficiente de correlación parcial tomando la raíz cuadrada de R^2 nos da como resultado r= +0.9786082. Esto significa que existe un alto grado de asociación entre las dos variables y además dicha relación es directa, dado el signo positivo, por lo que si una crece la otra también lo hará y viceversa. Tomamos el resultado positivo, ya que el signo obtenido en la ecuación de regresión del estimador de la variable independiente, que son las exportaciones, es positivo. Lo anterior puede corroborarse obteniendo la matriz de correlaciones, con esto se podrá verificar los cálculos previamente hechos (ver Cuadro 2.3).

Cuadro 2.3. Matriz de correlación.

	Correlation Matrix		
	X	PIB	
X	1.000000	0.978608	
PIB	0.978608	1.000000	

Obtención de los valores estimados de $\hat{Y}$

Finalmente, por lo que respecta a los valores estimados del Producto Interno Bruto a través de la ecuación de regresión ($\hat{Y}$), estos se obtienen al sustituir los valores que asume la variable independiente, las exportaciones, en dicha ecuación. Así siguiendo el ejemplo 1, tomáremos el valor que asumen en este caso las exportaciones en el primer y segundo trimestre de 1980 que equivalen a 5,810.920 y 6,059.96 millones de dólares, respectivamente. Estos valores se sustituyen en la ecuación de regresión, obteniendo para el primer trimestre de 1980:

$$P\hat{I}B_{1980Q1} = 4531969112 + 55871.51321 * \mathrm{X}$$
$$P\hat{I}B_{1980Q1} = 4531969112 + 55871.51321 * 5{,}810.92$$
$$P\hat{I}B_{1980Q1} = 4{,}856{,}634{,}005.54$$

Para el segundo trimestre de 1980:

$$P\hat{I}B_{1980Q2} = 4531969112 + 55871.51321 * \mathrm{X}$$
$$P\hat{I}B_{1980Q2} = 4531969112 + 55871.51321 * 6{,}059.96$$
$$P\hat{I}B_{1980Q2} = 4{,}870{,}548{,}247.19$$

Obtención de los residuos.

Los errores o residuos como se recordará se obtienen de restar al valor real de la variable dependiente, en este caso el PIB, el valor estimado de la misma mediante la ecuación de regresión, en este caso $P\hat{I}B$, es decir:

$$e_i = Y - \hat{Y} = PIB - P\hat{I}B$$

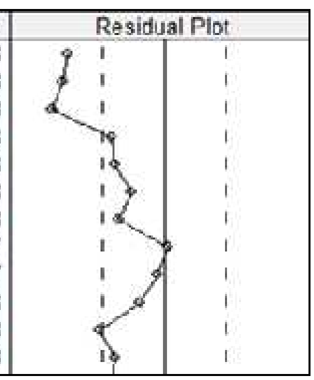
Así, el error tanto para el primero como para el segundo trimestre de 1980 es:

$$e_{1980Q01} = PIB - P\hat{I}B = 4{,}384{,}924{,}682.44 - 4{,}856{,}634{,}005.54 = \text{-}471709323$$
$$e_{1980Q02} = PIB - P\hat{I}B = 4{,}372{,}424{,}964.63 - 4870548247.19 = \text{-}498123283$$

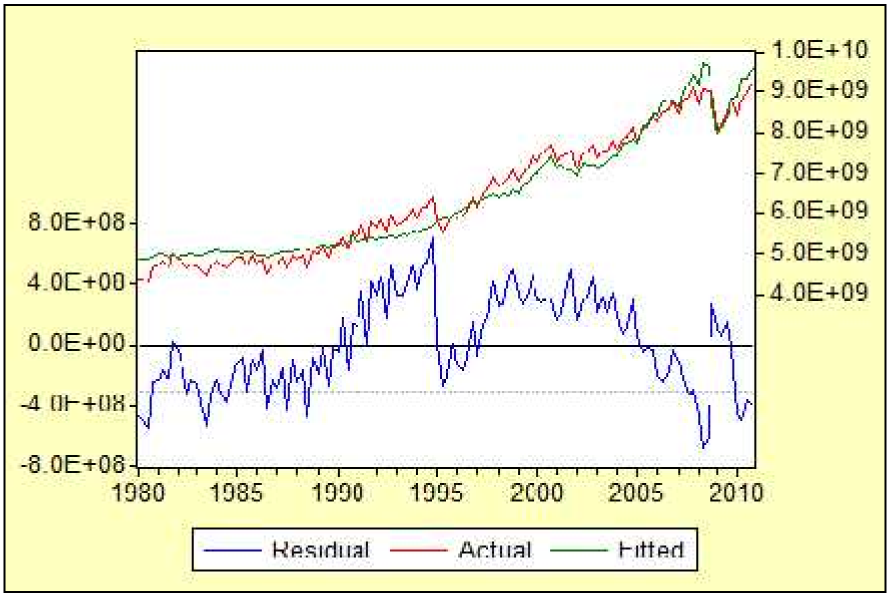
Mediante EViews se pueden obtener la tabla y la grafica de los valores reales del PIB, de los valores estimados de la variable dependiente $P\hat{I}B$, y los errores, esto véase en la obtención en el Cuadro 2.4 y 2.5 siguientes:

Cuadro 2.4. Valores reales y estimados de la variable dependiente y errores.

obs	Actual	Fitted	Residual
1980Q1	4.4E+09	4.9E+09	-4.7E+08
1980Q2	4.4E+09	4.9E+09	-5.0E+08
1980Q3	4.3E+09	4.9E+09	-5.6E+08
1980Q4	4.7E+09	4.9E+09	-2.5E+08
1981Q1	4.7E+09	5.0E+09	-2.4E+08
1981Q2	4.8E+09	5.0E+09	-1.7E+08
1981Q3	4.7E+09	4.9E+09	-2.3E+08
1981Q4	5.0E+09	5.0E+09	1.9E+07
1982Q1	4.9E+09	4.9E+09	-3.3E+07
1982Q2	4.8E+09	5.0E+09	-1.2E+08
1982Q3	4.7E+09	5.0E+09	-3.2E+08
1982Q4	4.8E+09	5.0E+09	-2.4E+08



Gráfica 2.5. Representación de los valores observados y estimados de la variable dependiente y lo errores.



IV.12.1.3. Ejemplo 3: Teoría de la inversión.

Planteamiento de la teoría económica.

En este ejemplo se desarrollara la teoría de la inversión en función de la tasa de interés. Dicha teoría establece que las familias ofrecen mano de obra a las empresas y éstas a su vez ofrecen bienes y servicios para su consumo, por su parte el gobierno por medio de políticas controla la economía influyendo en el nivel general de gastos de consumo, de los gastos de inversión y de los gastos de gobierno. En este caso el gobierno puede subir las tasas de interés, lo que provocará que la inversión disminuya, pues las personas preferirán poner a rentar su dinero en el sector financiero y éste a su vez se verá afectado pues las empresas solicitarán menos prestamos porque les saldrá más caro pedir un préstamo. Así las cosas, el elevamiento de la tasa de interés impediría el crecimiento de la empresa, lo que podría, eventualmente, conllevar a la desaparición de la misma, ocasionando el incremento del desempleo.

Cuando las empresas creen que se producirá una recuperación económica en el futuro inmediato, comienzan a hacer planes para expandir sus plantas y aumentar la producción. Si las empresas (los empresarios) temen que empeore la situación económica se mostrarán reacias a invertir. La decisión que toma un empresario de invertir, es una decisión para ampliar la reserva de capital de la planta, los inventarios y el equipo para el proceso de producción. La cantidad que invierta se verá afectada por su optimismo respecto al volumen de ventas futuras y por el precio de la planta y el equipo que se requiera para la expansión. Asimismo, las empresas piden préstamos para comprar bienes de capital por lo que la tasa de interés es un determinante fundamental para ampliar los activos fijos de compra. Cuanto más alto es el tipo de interés de esos préstamos, menores son los beneficios que pueden esperar obtener las empresas pidiendo prestamos para comprar nuevas maquinas o edificios y por lo tanto menos estarán dispuestas a pedir préstamos y a invertir. En cambio, cuando los tipos de interés son más bajos, las empresas desean pedir más préstamos e invertir más.

En el ejercicio se utilizaran datos de la economía mexicana como lo es la Formación Bruta de Capital (Inversión) y CETES a 28 días (tasa de interés) para el periodo comprendido entre el primer trimestre de 1985 y el cuarto trimestre de 2010; la inversión esta medida en miles de pesos constantes de 2003 y la tasa de interés en unidades porcentuales de los CETES a 28 días.

Así, el modelo matemático que se plantea tiene la siguiente estructura:

Inversión = f (tasa de interés)

Donde la Inversión está en función de la tasa de interés, en este caso la Formación Bruta de Capital está en función de los CETES a 28 días. Para hacer menos extensa la explicación del modelo simplificaremos tasa de interés al nombre solo de interés para su fácil utilización en EViews.

Mientras que el modelo econométrico es el siguiente:

$$Inversión = \hat{a} - \hat{b} * \text{int } erés + \hat{e}$$

La relación teórica entre la Inversión y la tasa de interés es inversa, es decir, si la tasa de interés (CETES a 28 días) aumenta, se espera que la Inversión (Formación Bruta de Capital) disminuya y viceversa.

De acuerdo con la teoría $\hat{a}$ representa la inversión (formación Bruta de Capital) independiente de cambios en la tasa de interés ya que si la tasa de interés fuese igual a cero la inversión seria igual a "$\hat{a}$ ", es decir, independientemente de los aumentos en la tasa de interés los empresarios deciden el monto de su inversión, sin embargo, al estar en ellos la decisión de la inversión, esta puede ser nula, por ello esta será cero o mayor que cero, pues está en los empresarios la decisión de adquirir más capital fijo o no invertir:

$$\hat{a} \geq 0$$

Por su parte $\hat{b}$ representa al multiplicador de la Inversión, misma que mide cuánto disminuye ésta al incrementarse la tasa de interés en una unidad porcentual adicional, por tanto, el valor que asume $\hat{b}$ <u>será menor que cero</u> por la relación inversa de las variables, la tendencia de la regresión será negativa, ya que, como lo habíamos afirmado, cuando aumenta la tasa de interés disminuye la inversión. Por cada unidad porcentual que se incremente la tasa de interés, dicho incremento se verá reflejado en la inversión según la magnitud de “b”, por ser este el multiplicador.

$$0 > \hat{b}$$

Pasemos ahora a analizar el comportamiento de las series. En la gráfica 3.1 observamos la tendencia de nuestras variables; la inversión por su parte tiene tendencia creciente, al paso del tiempo tiene 3 caídas notables que se explicaran históricamente más adelante. La tasa de interés cuenta con una tendencia hacia la baja, hacia el año 2010 es constante y estable con las cifras más bajas en el periodo establecido para nuestro estudio, aunque se pueden observar claramente tres periodos con un aumento considerable de la tasa de interés, estos valores altos coinciden con la tendencia a la baja del interés.

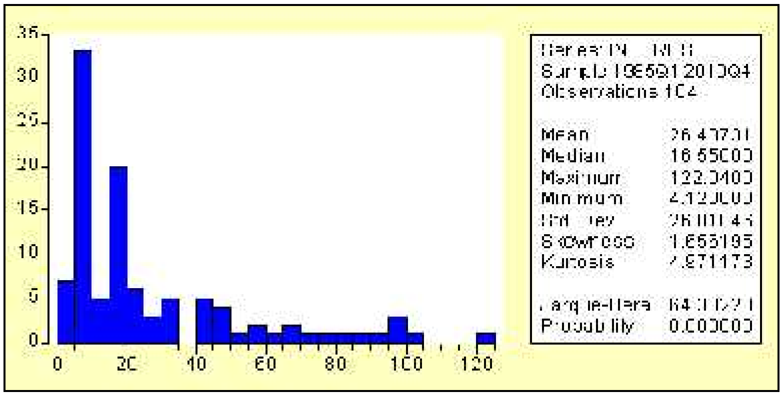
La inversión en el periodo que analizamos (1985-2010) es creciente, con algunas bajas; estas bajas pueden explicarse por los incrementos de las tasas de interés en tiempos de crisis, un claro ejemplo es el periodo comprendido entre 1982 y 1988, ya que ante la crisis, la economía se comienza a orientar hacia el mercado externo. En tiempos de crisis las tasas de interés aumentan para incrementar la entrada de capitales y reducir los efectos de la crisis, ante ello la inversión nacional cae como se observa en la gráfica 3.1. A partir de 1987 México crece constantemente de manera estable hasta el año de 1994, durante este periodo la inversión presentó un comportamiento creciente y la tasa de interés fue decreciente, sin embargo, ante la crisis de 1994-1995 la inversión se desploma y las tasas de interés se vuelven a incrementar. Al recuperarse de esta crisis México vuelve a tener un ritmo de crecimiento estable llegando a tener tasas de crecimiento hasta del 5% la cuales se vieron realmente afectadas al llegar el año 2008 con el inicio de la crisis económica mundial, llamando la atención que la inversión se desploma en 2008-2009 y que la tasa de interés no se incrementó.

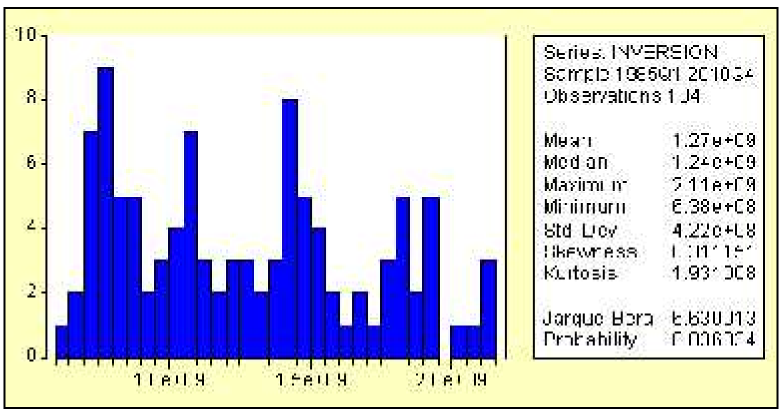
Gráfica 3.1. Gráficas individuales de las series.

Análisis estadístico inicial.

Para analizar la inversión y la tasa de interés del periodo especificado evaluaremos las cifras estadísticas obtenidas en el Cuadro 3.2.

Cuadro 3.1. Histogramas de las series.





Del análisis anterior se desprende que los datos son muy volátiles en ciertos puntos como lo vimos en la grafica de tendencia ya que los valores de la tasa de interés están entre 4.12 a 122.0400, medidos en unidades porcentuales, que comparados con la mediana del interés, ésta se encuentra cargada hacia la izquierda de la media con valor de 16.55; si hablamos de la inversión, pasa lo mismo, los datos están cargados a la izquierda de dicha media aritmética (1.27e).

Ninguna de las dos series, Inversión e Interés, son simétricas, pues la media no es igual a la mediana. La asimetría es positiva con valores en el interés de 1.65 y en la Inversión de 0.31, esta asimetría positiva quiere decir que la cola de la derecha es más larga que la de la izquierda, es decir, hay mas valores separados de la media a la derecha, por lo que las series no se distribuyen como una normal. Esto es más fácil de observar en los histogramas.

Con el análisis de la curtosis, en el caso del interés, ésta es leptocúrtica ya que está por encima de la normal (con valor 3 para la mesocúrtica), o sea que es más alta y fina. Hay una mayor centralización de las variables en torno a la media; esto se corrobora con el histograma donde tiene pocos valores pero muy altos. Con la inversión pasa lo contrario, al ser menor que 3 es platocúrtica, es decir, responde a una distribución por debajo de la normal. La no normalidad de las series se corrobora con la estadística Jarque Bera. Para ello se plantea la siguiente prueba de hipótesis:

Ho: La serie se distribuye como una normal.
Ha: La serie no se distribuye como una normal.

En términos ideales la estadística JB $\cong$ 6, por lo que, valores iguales o mayores a 6 en JB indican que la serie no se distribuye como una normal, en sentido contrario, valores en JB menores a 6 indican que la serie se distribuye como una normal. En los dos casos Ho se rechaza ya que el interés tiene un indicador de Jarque-Bera de 64.38 y la inversión de 6.63 afirmando que la serie no se distribuye como una normal

Asimismo, el programa proporciona una probabilidad asociada a la JB, misma que estableciendo α = 5%, indica la probabilidad mínima a la cual se rechaza la Ho. Así, si la probabilidad asociada a JB es menor a 0.05 se rechaza la Ho y si es mayor o igual a 0.05 no se rechaza dicha hipótesis. En este caso, se muestra la probabilidad para la inversión y la tasa de interés, misma que es menor a 0.05 rechazando la Ho y corroborando que las series no se distribuyen como una normal. La ausencia de normalidad en las series nos alerta sobre el posible incumplimiento de algunos supuestos del método de estimación al relacionarlas.

Identificación de la forma funcional.

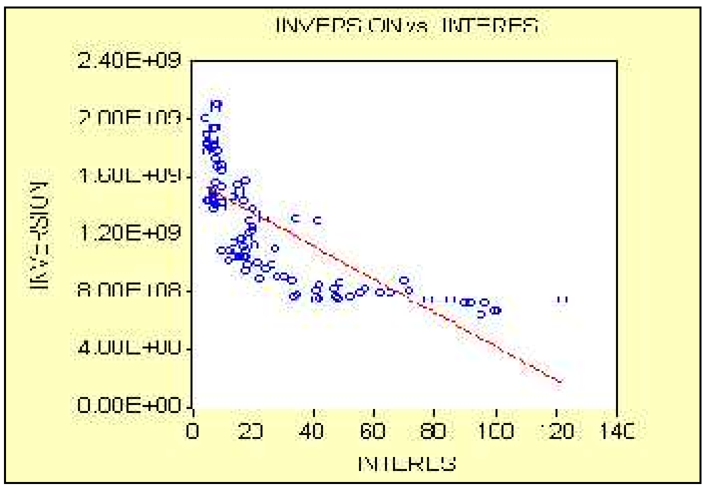
Una vez realizado el análisis gráfico y estadístico separado de cada variable, pasemos a determinar la forma funcional que corresponde a la relación entre la tasa de Interés con la Inversión a través del diagrama de dispersión que se muestra en la gráfica 3.2.

El mecanismo que utilizamos es la regresión simple lineal, pero podemos observar que en el diagrama de dispersión no se observa esta tendencia en las variables, en este caso es conveniente usar la regresión cuadrática para que se adapte a la curva que muestran los datos.

En el diagrama de “dispersión” se observa que no existe una relación lineal exacta entre las variables, ya que no toda la variación en la Inversión puede ser explicada por la tasa de interés. Si entre estas variables existiera una relación lineal perfecta, entonces todos o la mayoría de los puntos caerían a lo largo de la recta de regresión, que también ha sido trazada y que muestra la relación "promedio" que existe entre las dos variables.

En el diagrama se observa que los datos se concentran a la izquierda y del lado superior, es decir, en el periodo evaluado la mayoría de los datos obtenidos es una inversión alta con bajas tasas de interés, como observamos en el diagrama de dispersión la Inversión esta graficada del lado de las ordenadas por ello la concentración de los datos es notable en la parte superior indicando datos altos en la inversión y bajos en la tasa de interés ya que al estar graficada en las abscisas los datos se concentran del lado izquierdo.

Gráfica 3.2 Diagrama de dispersión con regresión.



Pasemos ahora a realizar la ecuación de regresión. Para generar la ecuación de regresión desde comandos, en el área de comandos escribimos: ls inversión c interés, seguido de oprimir la tecla enter, con lo que se desplegará una ventana como la contenida en el Cuadro 3.4.

Así, de acuerdo a los resultados de la estimación, la ecuación de regresión es:

$$Inversión = 1576792123 - 11608400.65 * Interes$$

Como se recordará, en teoría, $\hat{a}$debe ser mayor o igual que cero, pues representa la inversión que las empresas deciden hacer independientemente de la tasa interés. Si la tasa de interés fuera igual a 0% la inversión sería de 1576792123 miles de pesos constantes de 2003.

En cuanto al valor del estimador $\hat{b}$, observamos que la teoría se cumple con este dato ya que por cada unidad porcentual que se incremente el interés este se verá reflejado en una disminución de la inversión de 11608400.65 miles de pesos constantes de 2003, de forma inversa, si la tasa de interés disminuye en 1% entonces la inversión se incrementará en 11608400.65 miles de pesos constantes de 2003. La tasa de interés tiene relación inversa con la inversión, por ello, el signo del estimador es negativo. Con los resultados de la regresión corroboramos la teoría planteada en el principio del ejercicio.

Cuadro 3.2. Resultados de la ecuación de regresión.

Variable Independiente: Inversión
Variable Independiente: Interes
Método: Minimos Cuadrados
Periodo: 1985Q1 2010Q4
Observaciones Incluidas: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.58E+09	39547999	39.87034	0.0000
INTERES	-11608401	1051918.	-11.03547	0.0000

R-squared	0.544198	Mean dependent var	1.27E+09
Adjusted R-squared	0.539730	S.D. dependent var	4.22E+08
S.E. of regression	2.86E+08	Akaike info criterion	41.80148
Sum squared resid	8.36E+18	Schwarz criterion	41.85234
Log likelihood	-2171.677	F-statistic	121.7815
Durbin-Watson stat	0.218591	Prob(F-statistic)	0.000000

Verificación de la teoría económica.

A continuación procedemos a realizar la prueba de significación estadística de los parámetros poblacionales. Para ello establecemos la siguiente prueba de hipótesis:

Ho: $\beta = 0$, por tanto, los CETES a 28 días (Interés) no explican a la Inversión.
Ha: $\beta \neq 0$, por tanto, los CETES a 28 días (Interés) explican a la Inversión.

Para probar la significación estadística de los parámetros, en este caso el de la variable explicativa, es decir, la tasa de interés, se compara la "t" calculada del parámetro (t_β) con la "t_α", también conocida como "t" teórica o de tablas. Al respecto, la "t" calculada del parámetro β se obtiene de la siguiente manera:

$$t_{\hat{b}} = \frac{\hat{b} -}{s_{\hat{b}}} = \frac{-11608401 - 0}{1051918} = -11.0354619$$

El resultado obtenido de la "t_β" puede corroborarse en la ventana de los resultados de la ecuación de regresión (ver Cuadro 3.2). Dicho resultado se compara con la "t_α" con un nivel de significación estadística α=5% y grados de libertad igual a n-k=104-2=102, cuyo valor en tablas es igual a $t_\alpha = \pm 2.2749$.

Los criterios para rechazar o no la Ho son los siguientes:

1. Si t ≥ t_α, ambas en términos absolutos, entonces se rechaza la Ho.
2. Si t < t_α, ambas en términos absolutos, entonces no se rechaza la Ho.

Así, al ser t = 11.03547 > t_α = 2.2694 se rechaza la Ho, por tanto, los CETES a 28 días (Interés) explican a la Inversión.

En cuanto a la probabilidad asociada al estadístico "t" (p-value), la cual representa la probabilidad mínima a la cual se rechaza la Ho. Los criterios, con un nivel de significación al 5%, para rechazar o no la Ho con el p-value son:

1. Si la probabilidad asociada al estadístico "t" es menor o igual a 0.05 se rechaza la Ho.

2. Si la probabilidad asociada al estadístico "t" es mayor a 0.05 no se rechaza la Ho.

Por lo que en nuestro caso se corrobora que se rechaza la Ho ya que la probabilidad asociada a "t" es menor a 0.05.

En cuanto a la prueba de significación estadística de $\hat{a}$, se plantea la siguiente prueba de hipótesis:

Ho: α = 0, por tanto, su valor no tiene o no es de relevancia económica.
Ha: α > 0, por tanto, su valor si es de relevancia económica.

Como la probabilidad asociada a t_a es menor a 0.05 se rechaza la Ho, mostrando que $\hat{a}$ es significativo y por ende de relevancia económica.

Por lo que respecta a la prueba de bondad de ajuste, el coeficiente de determinación (R^2) es de 0.544198 lo que significa que el 54.4198% de los cambios o variaciones en la inversión se deben o explican por cambios o variaciones en la tasa de interés, y el resto se explica por variables que no fueron incluidas explícitamente en el modelo, esta puede ser una razón para que el R^2 sea tan bajo, ya que se deben de incluir mas variables para poder explicar la inversión; otro factor es la mala especificación de la forma funcional. El coeficiente de determinación ajustado ($\overline{R}^2$), el cual se ajusta mediante la introducción de grados de libertad, determina la variación que es explicada por la variable independiente, con respecto a la variable dependiente, cuando se introduce una variable adicional al modelo. En nuestro caso $\overline{R}^2$ es de 0.53973, lo que significa que el 53.97% de las variaciones en la inversión se explican por las variables independientes, en este caso la tasa de interés y una variable adicional. En cuanto al coeficiente de correlación parcial (r) este es igual a:

$$r = \pm\sqrt{R^2} = \pm\sqrt{0.544198} = -0.73769777$$

En teoría se considerará un "r" aceptable cuando éste asume un valor mayor o igual a 0.5 en términos absolutos, trabajando con las series en sus valores originales. Así, para nuestro ejemplo, obtuvimos un r= -0.73769777, lo que significa que existe un mediano grado de asociación entre las dos variables y que dicha relación es inversa, dado el signo negativo, por lo que si una crece la otra disminuirá y viceversa.

Finalmente, por lo que respecta a los valores estimados de la inversión a través de la ecuación de regresión ($\hat{Y}$), estos se obtienen al sustituir los valores que asume la variable independiente, la tasa de interés, en dicha ecuación. Así, por ejemplo, el valor que asume la tasa de interés en el primer y segundo trimestre de 1985 es 55.29% y 61.98% respectivamente. Estos valores se sustituyen en la ecuación de regresión, obteniendo para el primer trimestre de 1985:

$$Inver\hat{s}ión_{1985Q01} = 1576792123 - 11608400.65 * Interes$$
$$Inver\hat{s}ión_{1985Q01} = 1576792123 - 11608400.65(55.29)$$
$$Inver\hat{s}ión_{1985Q01} = 934963651$$

Para el segundo trimestre de 1985:

$$Inver\hat{s}ión_{1985Q02} = 1576792123 - 11608400.65 * PIB$$
$$Inver\hat{s}ión_{1985Q02} = 1576792123 - 11608400.65(61.98)$$
$$Inver\hat{s}ión_{1985Q02} = 857303451$$

Los errores o residuos se obtienen de restar al valor real de la variable dependiente, en este caso la inversión, el valor estimado de la misma mediante la ecuación de regresión, en este caso $Inver\hat{s}ión$, es decir:

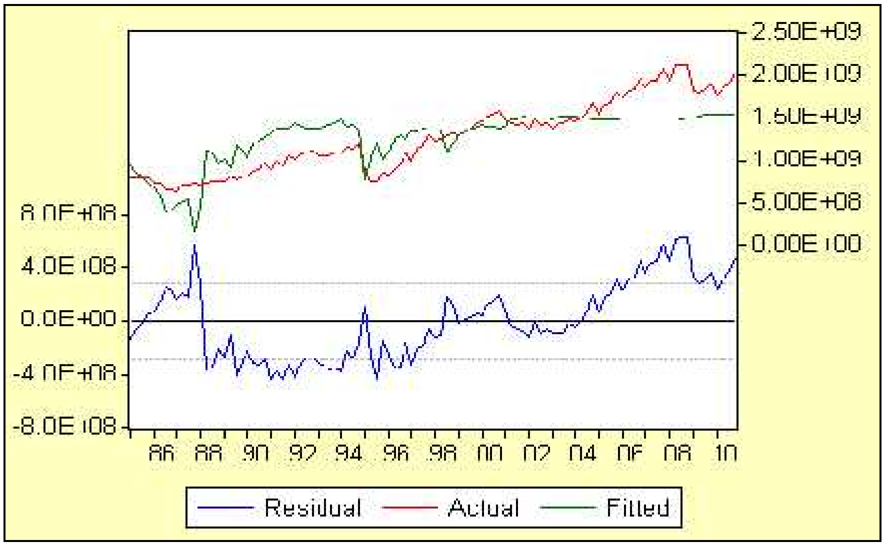
$$e_i = Y - \hat{Y} = Inversión - Inver\hat{s}ión$$

Así, el error tanto para el primero como para el segundo trimestre de 1985 es:

$$e_{1985Q01} = Inversión - Inver\hat{s}ión = 798123462.64 - 934963651 = \text{-}136840188.42$$
$$e_{1985Q02} = Inversión - Inver\hat{s}ión = 793779933.59 - 857303451 = \text{-}63523517.12$$

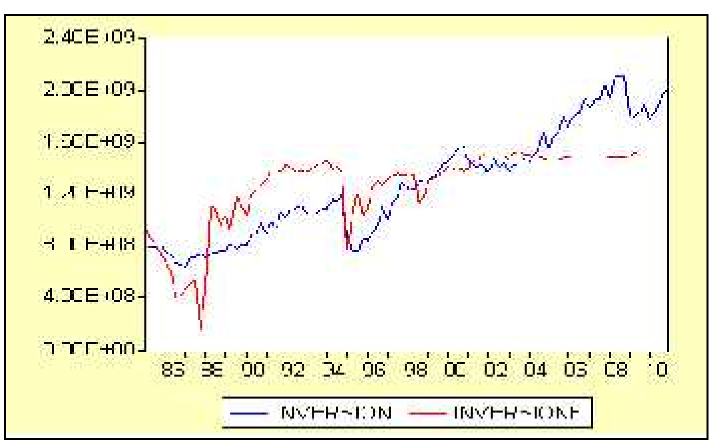
Estos datos también se representan en la grafica de residuos, esta muestra los datos reales (Actual), la inversión estimada (Fitted) y los errores (Residual) la grafica de residuos. En la gráfica 3.3 observamos como los errores son notables en 1988 a 1994 dadas las situaciones que analizábamos antes, ya que en tiempos de crisis, las variables pueden comportarse de manera distinta o atípica; de la grafica también se desprende la magnitud de los errores.

Gráfica 3.3 Representación de los valores observados y estimados de la variable dependiente y lo errores.



En la comparación de la inversión real y la estimada en el Cuadro 41 se observa que se separan en ciertos periodos notablemente. Nuestro análisis de regresión muestra la relación de dos variables, de acuerdo a la experiencia; en este caso, dadas las cifras analizadas anteriormente podemos concluir que éstas de acuerdo a los datos están relacionadas, pero nuestros supuestos de teoría no se aplican en los tiempos de crisis, al no actuar las variables según la teoría, se muestra en la variable de inversión estimada, un gran cambio, que aunque conserva su tendencia, elimina los valores más altos. Ese resultado también puede deberse a la mala especificación del modelo y a la incorrecta forma funcional.

Gráfica 3.4 Pronóstico de la variable dependiente.



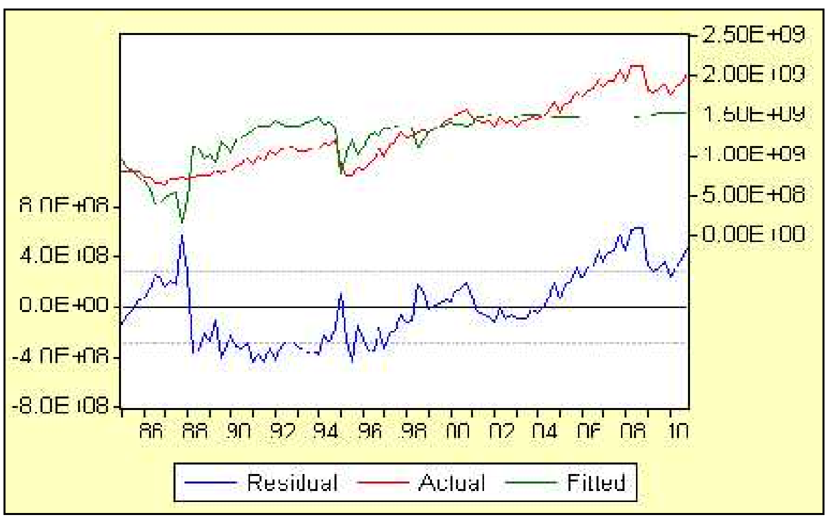
En el Cuadro 14 se observa la inversión en datos reales en color azul, y la inversión resultante de nuestra regresión en color rojo, las dos series cambian mucho, mientras en una vemos una tendencia marcada, en la serie estimada Inversionf tenemos que es mas horizontal con algunas caídas notables, pero dentro de un rango mas pequeño, esto puede deberse como ya lo habíamos dicho a una mala especificación del modelo o a el manejo de una forma funcional equivocada.

$$e_{1985Q01} = Inversión - \hat{Inversión} = 798123462.64 - 934963651 = -136840188.42$$

$$e_{1985Q02} = Inversión - \hat{Inversión} = 79377993359 - 857303451 = -63523517.12$$

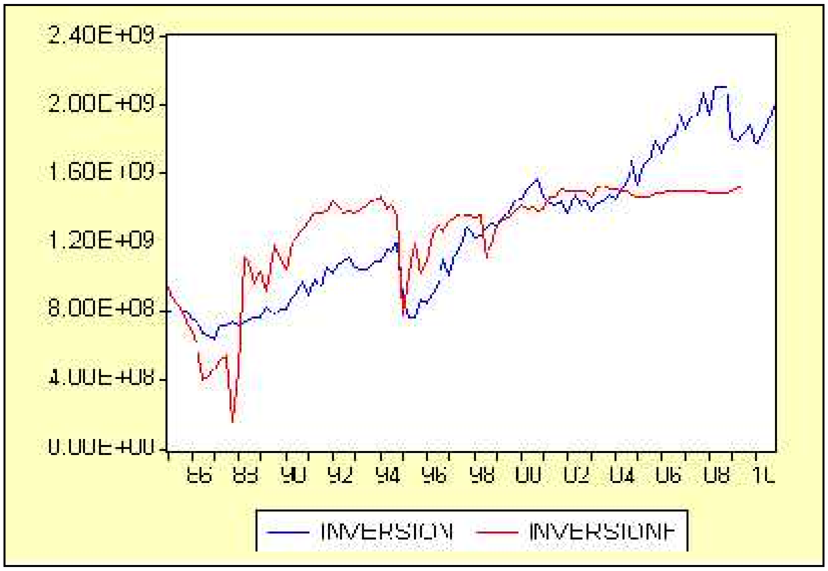
Estos datos también se representan en la grafica de residuos, esta muestra los datos reales (Actual), la inversión estimada (Fitted) y los errores (Residual) la grafica de residuos. En el Cuadro 40 observamos como los errores son notables en 1988 a 1994 dadas las situaciones que analizábamos antes, ya que en tiempos de crisis, las variables pueden comportarse de manera distinta o atípica; de la grafica también se desprende la magnitud de los errores.

Gráfica 3.5 Representación de los valores observados y estimados de la variable dependiente y lo errores.



En tanto que dentro del Cuadro 3.7 se procede hacer la comparación de la inversión real y la estimada, se observa que se separan en ciertos periodos notablemente.

Gráfica 3.6 Pronóstico de la variable dependiente.



Nuestro análisis de regresión muestra la relación de dos variables, de acuerdo a la experiencia; en este caso, dadas las cifras analizadas anteriormente podemos concluir que éstas de acuerdo a los datos están relacionadas, pero nuestros supuestos de teoría no se aplican en los tiempos de crisis, al no actuar las variables según la teoría, se muestra en la variable de inversión estimada, un gran cambio, que aunque conserva su tendencia, elimina los valores más altos. Ese resultado también puede deberse a la mala especificación del modelo y a la incorrecta forma funcional.

IV.13.- Reactivos para reafirmar los conocimientos.

A.- Primer ejercicio de preguntas y respuestas:

1.- Indique que significa y para qué sirve:

a) El análisis de regresión simple;

Se usa para probar hipótesis sobre la relación entre una variable dependiente Y, y una variable independiente, X, y para pronóstico, que contrasta con el análisis de regresión múltiple en el que hay más de una variable independiente.

b) El análisis de regresión lineal;

Supone que hay una relación lineal entre X y Y.

c) El diagrama de dispersión;

Su fin es determinar por inspección ocular la relación funcional entre X y Y.

2.- Establezca la relación general entre el consumo Y, y el ingreso disponible X, con:

a) Forma lineal exacta o deterministica: corresponde a un modelo económico;

$\hat{Y}_i = a + bX_i$

b) Forma estocástica: corresponde a un modelo econométrico;

$\hat{Y}_i = a + bX_i + e_i$

c) ¿Por qué se supone que la mayoría de los valores observados de Y no caen exactamente sobre una línea recta constuida con sus valores estimados?

d)

Porqué hay otras variables explicativas omitidas que también tienen efecto en Y; por errores de medición de Y_i y por la aleatoriedad inherente a la conducta humana.

3.- Con el método de mínimos cuadrados ordinarios,MCO, para estimar los valores poblacionales de ~ y con los valores muestrales de a y b diga:

a) ¿Por qué se usa éste método?

Da la línea recta óptima que minimiza las diferencias de Y_i con $\hat{Y}_i$

b) ¿Qué *propiedades* tienen los estimadores a y b?

El MCO produce estimadores con propiedades de insesgabilidad, eficiencia, suficiencia y consistencia.

c) ¿Por qué se miden verticalmente las desviaciones?

Porque tratamos de explicar movimientos en Y_i que se miden en el eje vertical.

d) ¿Por qué no simplemente se toman la suma de las desviaciones sin elevarlas al cuadrado?

Porque su suma es igual cero.

e) ¿Porqué no tomarse la suma de las desviaciones absolutas?

Porque su suma no es un mínimo.

4.- i) Establezca la diferencia entre a y b por un lado, y ~ y   por el otro; en una población *infinita*, ~ y   son los parámetros de la línea de regresión verdadera pero desconocida en tanto que a y b son sus estimadores o parámetros de la línea de regresión estimada conocida.

ii) ¿Establezca la diferencia entre U_i y e_i, es decir ¿qué significa U_i y e_i?

U_i es el término de perturbación aleatorio, error estocástico en la relación verdadera pero desconocida en una población infinita entre X_i y Y_i, mientras que $e_i = Y_i - \hat{Y}_i$ en una muestra seleccionada aleatoriamente.

iii) Escriba las ecuaciones para las relaciones verdaderas en el universo y estimadas con la muestra entre X e Y.

Para la verdadera: $Y_i = \sim + \quad X_i + U_i$

Para la estimada: $\hat{Y}_i = a + bX_i + e_i$

5.- Con los siguientes datos: a) hallar los valores de a y b; b) representar gráficamente la línea de regresión y c).- mostrar las desviaciones de cada Y_i del correspondiente $\hat{Y}_i$. Interprételas.

a) Para hallar a y b tenemos:

Año	Y=Consumo	X=Ingreso	X_iY_i	X_i^2
1	102	114	11,628	12,996
2	106	118	12,508	13,924
3	108	126	13,608	15,876
4	110	130	14,300	16,900
5	122	136	16,592	18,496
6	124	140	17,360	19,600
7	128	148	18,944	21,904
8	130	156	20,280	24,336
9	142	160	22,720	25,600
10	148	164	24,272	26,896
11	150	170	25,500	28,900
12	154	178	27,412	31,684
Suma	1,524	1,740	225,124	257,112
Media	127	145		

Se calculan:

$n=12, \sum Y_i = 1{,}524, \overline{Y} = 127, \sum X_i = 1{,}740, \overline{X} = 145, \sum X_i Y_i = 225{,}124, \sum X_i^2 = 257{,}12$

$$b = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} = \frac{(12)(225{,}14) - 1{,}740)(1{,}524)}{(12)(257{,}112) - (1{,}740)^2} = \frac{49{,}728}{57{,}744} = 0.86$$

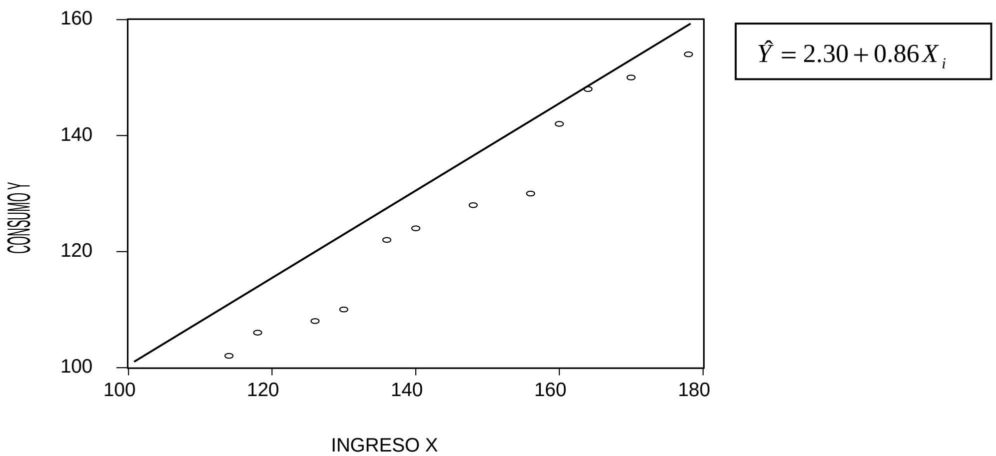
$a = \overline{Y}_i - b\overline{X}_i$ =127-(0.86*145)=127-124.70=2.30

Luego la ecuación de regresión: $\hat{Y}_i = 2.30 + 0.86X_i$

b) Para ello obtenemos 2 puntos cualesquiera sobre la línea de regresión, digamos

Cuando Xi = 114, $\hat{Y}_i$ =2.30+(0.86*114)=100.34

Cuando Xi = 178, $\hat{Y}_i$ =2.30+(0.86*178)=155.38



B.- Segundo ejercicio de preguntas y respuestas:

1. ¿Qué es regresión?

   Es expresar con el método MCO la relación funcional o la dependencia de una variable (Y) con respecto a otra (X).

2. ¿Qué es correlación?

Es la cuantificación por medio del coeficiente de correlación (r), la relación o el grado de relación entre dos variables, es decir, X e Y.

3. ¿Cuál es la diferencia entre la correlación y regresión simple y múltiple?

La regresión y la correlación es simple cuando se utilizan dos variables y es múltiple cuando se utilizan más de dos.

4. ¿Entre que valores oscila el valor de r?

El valor de r oscila entre los valores –1 hasta + 1, pasando por el cero. En donde si r tiene un valor de – 1, decimos que existe una fuerte relación o correlación negativa o inversa entre X e Y, es decir, tiene una relación decreciente o lo que es lo mismo, a medida que X aumenta Y disminuye. Si r toma el valor de + 1, entonces decimos que existe una fuerte relación positiva entre X e Y, es decir, tiene una relación creciente, lo cual significa que a medida que X aumenta Y también lo hace. Por último, si r tiene un valor de 0, se sostiene que existe nula relación o correlación entre X e Y.

5. ¿Para qué sirve el diagrama de dispersión?

Para determinar la forma funcional adecuada que exprese la relación entre X e Y. Es decir, el diagrama de dispersión nos ayuda a determinar si utilizamos, digamos, la forma funcional de la línea recta, la parábola, la exponencial, la logarítmica, etc.

6. En Y = f (X), escriba las ecuaciones normales necesarias para hallar los valores de los parámetros en una recta.

$$\sum Y_i = na + b\sum X_i \quad (1)$$

$$\sum X_i Y_i = a\sum X_i + b\sum X_i^2 \quad (2)$$

7. En Y = f (X), de una parábola escriba las ecuaciones normales necesarias para hallar los valores de los parámetros.

$$\sum \boldsymbol{y} = \boldsymbol{na} + \boldsymbol{b}\sum \boldsymbol{x} + \boldsymbol{c}\sum \boldsymbol{x}^2 \qquad (1)$$

$$\sum \boldsymbol{xy} = \boldsymbol{a}\sum \boldsymbol{x} + \boldsymbol{b}\sum \boldsymbol{x}^2 + \boldsymbol{c}\sum \boldsymbol{x}^3 \qquad (2)$$

$$\sum \boldsymbol{x}^2\boldsymbol{y} = \boldsymbol{a}\sum \boldsymbol{x}^2 + \boldsymbol{b}\sum \boldsymbol{x}^3 + \boldsymbol{c}\sum \boldsymbol{x}^4 \qquad (3)$$

8. ¿Para qué sirve el error estándar?

En econometría se usan dos: El de la estimación que hacen los puntos de la de la *recta de regresión de* los puntos observados o reales de la variable dependiente y, el del coeficiente de la *variable* *independiente*, también llamado el error estándar del estimador. El primero indica qué tan próximos están los valores estimados con la ecuación de regresión, de los valores reales u observados de la variable dependiente; el segundo sirve para calificar la calidad de la estimación del parámetro del universo con el coeficiente o "estadístico muestral" de la variable independiente. De tal suerte que, en este segundo caso, si se obtiene una error estándar de la estimación pequeño, se dice que el valor del parámetro muestral es cercano al poblacional, y por tanto, la estimación es adecuada y viceversa.

El error estándar del coeficiente de la variable independiente también sirve para probar hipótesis y para obtener el intervalo de confianza en que se estima que esté el valor del parámetro poblacional, con determinada probabilidad asociada al nivel de significación.

9. ¿Cuál es la relación de una pendiente positiva con r?

Tiene una relación directamente proporcional, es decir, que si la pendiente es positiva el coeficiente de correlación (r) también lo es, lo cual muestra que existe una relación directa entre X e Y.

10. ¿Cuál es la relación de una pendiente negativa con r?

Tiene una relación inversamente proporcional, es decir, que si la pendiente es negativa el coeficiente de correlación (r) también lo es, lo cual muestra que existe una relación inversa entre X e Y.

Ejercicios para responder fuera del aula:

11.- ¿Qué tienen en común y en que difieren:

a).- Las estadísticas **t** y **F;**

b).- El error estándar de la estimación con el error estándar del coeficiente de la variable independiente, con la suma de cuadrados de los errores y con la desviación estándar de la variable dependiente?;

c).- ¿Los coeficientes: de determinación y el ajustado? ¿ Cuál de ellos es más exacto, porqué?;

d).- En una regresión *múltiple*, ¿son mayores los valores de los coeficientes de correlación, de determinación y el ajustado, que sus valores obtenidos en una regresión *simple*? ¿ por qué?

Preparación de exámenes sobre regresión y correlación simple:

C.- EJERCICIO CON RESPUESTAS.

Nombre del alumno:____
__

Planteamiento: La teoría económica establece que el consumo, Y_i, es función del ingreso, X_i, relación que en los últimos años es:

Años	Y_i	X_i
1	102	114
2	100	118
3	108	126
4	110	130
5	122	136
6	124	140
7	128	148
8	130	156
9	142	160
10	148	164
11	150	170
12	154	178

Como se vio en la Practica I, punto 5, se estima la regresión y se hallan los siguientes resultados:
$\hat{Y}_i = 2.30 + 0.86X_i$

donde:

$\hat{a} = 2.30$, $\hat{b} = 0.86$

$S_{\hat{a}} = 7.17$, $S_{\hat{b}} = 0.05$

Preguntas:

1. ¿Cuál es el significado de a y de b?

El estimador $\hat{a}$=2.30 es la ordenada al origen y el valor del consumo total cuando el ingreso disponible es cero. Como $\hat{a}$>0 se confirma que siempre habrá un consumo básico. El estimador $\hat{b} = \frac{dY}{dX} = 0.86$ es la pendiente de la línea de regresión estimada. Mide la propensión marginal al consumo o el cambio que experimenta el consumo con el cambio en una unidad adicional en el ingreso. Como $0<\hat{b}<1$ confirma la teoría del consumo de que las personas incrementan sus gastos de consumo cuando aumenta el ingreso disponible, pero en menor proporción que este.

2. Sabiendo que $¦ = \hat{b}\frac{\overline{X}}{\overline{Y}}$

Donde:

¦ = Elasticidad ingreso del consumo

$\overline{X} = 145$

$\overline{Y} = 127$

Calcule e interprete ¦

Tenemos $¦ = \frac{\overline{X}}{\overline{Y}}\hat{b} = \frac{145}{127} * 0.86 = 0.98$

Interpretación: ¦ mide el cambio porcentual en consumo derivado de un cambio porcentual en el ingreso disponible.

3. Describa la hipótesis nula y alternativa para probar la significación estadística de los parámetros de la ecuación de regresión estimada.

Para probar la significación estadística de los parámetros poblacionales ~ y con los muestrales a y b establecemos:

Hipótesis nula: H_o: ~ =0 versus H_A: ~ ≠ 0

H_o: =0 versus H_A: ≠ 0

Al investigador le interesa rechazar H_o y aceptar H_A, es decir que ~ y ≠0 con una prueba de dos colas, por lo general salvo planteamientos específicos.

4. ¿Cuál es la forma de la distribución muestral de $\hat{a}$ y $\hat{b}$?

Como se supone que Ui tiene una distribución normal, Yi también tiene una distribución normal (dado que se supone que Xi es fija). Como resultado a y b también tienen distribución normal

5. ¿Qué distribución se usa para probar la significación estadística de ~ y ?.

Se establece usar la distribución t de student porque ~ y tienen distribución normal, pero se desconocen, y n<30.

6. ¿Para que sirven los grados de libertad y cuántos son?

Los grados de libertad sirven para mejorar la exactitud o bondad de ajuste de los estimadores $\hat{a}$ y $\hat{b}$. Se calculan con ***n-k*** =12-2=10

7. Con ~ =5% pruebe si ~ y son significativos estadísticamente. Interprete los resultados.

$t_a = \frac{\hat{a}-a}{S_{\hat{a}}} = \frac{2.30-0}{7.17} = 0.32$. Dado que $t_{\sim} = \pm 2.228$ en una prueba de dos colas decimos que no es significativa estadísticamente: no rechazamos Ho.

Igualmente si $t \ = \frac{\hat{b}-}{S_{\hat{b}}} = \frac{0.86-0}{0.05} = 17.2$. Como $t_{\beta} > t_a$ decidimos aceptar H_A y decimos que <u>estadísticamente difiere significativamente de cero</u>, que X_i si explica a Y_i. Observación si hubiéramos aceptado H_o, no se cumpliría la relación hipotética de Y_i con X_i, mismo que debemos modificar y reestimar hasta lograr una relación estimada de consumo satisfactoria.

8. Construya la banda de confianza al 95% para ~ y .

La banda o intervalo de confianza se construye así:

$\sim = \hat{a} \pm t_{\sim} S_a = 2.30 \pm 2.228(7.17) = 2.30 \pm 15.97$

Tal que ~ está entre –13.67 y 18.27 con confianza del 95%. Como dice Salvatore (1991:103) el intervalo es más amplio (e insensato), lo cual es consecuencia de que es <u>altamente insignificativa.</u>

Para tenemos $= \hat{b} \pm t_{\sim} S_b = 0.86 \pm 2.228(0.05) = 0.860 \pm 0.11$, decimos que está entre 0.75 y 0.97, en otras palabras 0.75< <0.97 con 95% de confianza. Al no incluir este intervalo de confianza el cero podemos decir que X_i si explica a Y_i.

9. ¿Cuántos usos tiene el error estándar de $S_{\hat{a}}$ y $S_{\hat{b}}$?

Aquí son 2: probar hipótesis y para la banda de confianza pero también sirve para obtener matemáticamente el tamaño de la muestra.

Encontrar R^2 con cualquiera de los métodos conocidos e interprételo.

Tenemos R^2=0.9687 ó 96.87% indica que la variable dependiente es explicada en un 96.87% por las variables independientes.

D.PRACTICA A RESOLVER: REGRESIÓN Y CORRECCIÓN SIMPLE.

1. ¿Por qué es mejor el método de mínimos cuadrados para estimar parámetros muestrales de los parámetros de la población?
2. Explique las propiedades de los estimadores obtenidos con el método de mínimos cuadrados.
3. ¿Qué es un estimador insesgado?
4. ¿Qué es un estimador consistente?

5. ¿Qué es un estimador suficiente?
6. ¿Qué es un estimador eficiente?
7. Explique usted las hipótesis o supuestos básicos del modelo de regresión lineal clásico.
8. Con los datos de Y_i: inversión en miles de pesos, X_i: tasa de interés en porcientos, estime la ecuación de regresión de Y_i sobre X_i.

Y_i Miles de pesos	X_i Tasa de interés
6	9
8	10
8	8
7	7
7	10
12	4
9	5
8	5
9	6
10	8
10	7
11	4
9	9
10	5
11	8

9. Con los datos anteriores probar con un nivel de significación del 5%, la significación estadística de los parámetros $\sim$ y .
10. Hallar e interpretar R^2 y $\overline{R}^2$

E.- Primer examen parcial: Conocimientos básicos.

I.En las siguientes diez frases escriba en las celdas vacías correspondientes **No** cuando sean falsas y **SI** cuando sean verdaderas:

1.- La econometría expresa y mide las relaciones entre Ye X:

Si______ No______

2.- Los modelos multiecuacionales son los únicos que se usan:

Si______ No______

3.-La covarianza mide la relación conjunta de Y e X:

Si______ No______

4.- El valor del coeficiente de determinación está entre -1,0 y +1:

Si______ No______

5.-La variable dicotómica expresa cuantitativamente un atributo:

Si______ No______

6.- Una variable retardada es estática en el tiempo:

Si______ No______

7.-La variable aleatoria también se llama estocástica:

Si______ No______

8.- Según los supuestos básicos del Modelo Lineal Simple "todas las perturbaciones aleatorias tienen la misma varianza":

Si______ No______

9.- El método de mínimos cuadrados no produce la variación mínima en los estimadores con respecto a los parámetros que estiman:

Si______ No______

10.-El diagrama de dispersión ayuda a identificar la forma funcional:

Si______ No______

II.- Suponga usted concretamente que el ñconsumo (**Y_i**) y el ingreso (**X**i) para los últimos 4 años (en miles de millones de pesos) son los siguientes

Año	***(Y_i)***	***(X_1)***				
1	3	5				
2	4	6				
3	5	8				
4	8	9				

1.- Encuentre la ecuación de regresión e interprete estadística y económicamente sus coeficientes; 2. Calcule e interprete **r** e **R^2**; 3.- a). Calcule el error estándar de *estimación* o de regresión: b). *compárelo* con el error estándar del coeficiente de regresión de la variable independiente; c). *interprételo* para conocer la bondad en el ajuste de los puntos *estimados* que constituyen la recta de regresión con respecto a los valores *reales* u observados de Y; 4.-*Pruebe estadísticamente la hipótesis*de que el consumo en México depende de las variaciones que experimenta el ingreso con α=10% e interprete *económicamente* los resultados.

Como con el MLS se ha venido hablando de relaciones lineales entre las variables, es oportuno decir que éstas en la vida real no siempre se presentan, por lo que es oportuno abordar el tema de las relaciones no lineales como se indica en el siguiente apartado.

IV.14.- Transformaciones de formas funcionales no lineales en lineales.

Propósito.

En economía existen fenómenos (variables) que no siempre tienen un comportamiento lineal, digamos como el comportamiento del valor de las acciones de cualquier empresa en el mercado bursátil. Este comportamiento se **detecta** al graficar los valores del fenómeno y obtener un diagrama de dispersión en forma de “sierras”; también ello se comprueba cuando cierta teoría económica así lo establece, digamos el valor de la producción en función de los insumos de mano de obra y de capital.

Comenta el profesor Salvatore (1991, p. 136) que la teoría económica puede a veces sugerir la forma funcional de una relación económica; también, que la dispersión de los valores observados puede sugerir la forma funcional apropiada en una relación de dos variables, y que cuando ninguna teoría ni dispersión de puntos es de ayuda, la función lineal se trata usualmente primero debido a la simplicidad.

Algunas de las transformaciones de funciones no lineales a lineales más útiles y comunes son las funciones logaritmo doble o doble-log, logarítmica – lineal, lineal-logarítmica, recíproca, y la polinomial. Una de las ventajas de la forma doble-log es que las pendientes representan elasticidades. La función semilog es apropiada cuando la variable dependiente crece en el tiempo a un ritmo relativamente constante, como en el caso de la fuerza laboral y de la población. Las funciones recíprocas y polinomial son apropiadas para estimar curvas de costo medio y costo total.

Al respecto, conviene señalar que la estimación de una función doble-log transformada por el método de MCO arroja estimadores de pendiente insesgados. Sin embargo b_0= su antilog $b_{0,}$ es decir, el antilogaritmo del logaritmo de b_0 es un estimador sesgado pero consistente de b_0. El hecho de que b_0 sea sesgado no es de mucha importancia, porque la constante no muestra un interés especial en economía. En las otras funciones transformadas b_0 también es insesgado.

Se recomienda trasformar estas “formas funcionales” no lineales en lineales cuando se aplica el análisis de regresión con el objetivo de que al utilizar el método de mínimos cuadrados, se obtengan funciones lineales trasformadas *cuyas pendientes sean estimadores insesgados*. En este contexto conviene recordar (ver supuestos) que el método de mínimos cuadrados ordinarios, MCO, se aplica principalmente al análisis de modelos con relaciones lineales en los parámetros; de lo anterior se infiere que a veces los modelos pueden no ser lineales en las variables pero lineales en los parámetros y por consiguiente pueden ser trabajados con MCO, siempre y cuando dichas variables sean transformadas en lineales.

Beneficios de las transformaciones:

1.- Estas *"nuevas" pendientes* aparte de ser estimadores lineales en los logaritmos de las variables Y e X, son insesgadas y su valor es igual a la **elasticidad** de la variable dependiente con respecto a la independiente o explicativa, como es el caso de la forma funcional *doble logarítmica*, también expresada por motivos prácticos **log log** o de elasticidad constante;

2.- Con la forma funcional **log lin** la pendiente de la variable independiente, X, también es lineal y mide una **tasa de crecimiento,** es decir, mide el cambio porcentual o relativo constante en la variable dependiente, Y, provocado por el cambio unitario absoluto que experimente la variable independiente, que debe ser la *variable tiempo* ;

3.- Con la forma funcional **lin log** la pendiente de la variable independiente, X, mide el cambio absoluto en la variables dependiente, Y, originado en un cambio porcentual o relativo en X;

4.- El modelo o forma funcional recíproco (en X) aun cuando no es lineal (Gujarati, 1990) en X puesto que entra inversamente o de manera reciproca, es lineal en sus parámetros y en consecuencia, por esa razón es un "modelo de regresión lineal". Su connotación principal es que a medida que X aumenta indefinidamente, su coeficiente se acerca a cero y la variable dependiente Y se aproxima al valor límite o asintótico, que es el valor de la ordenada al origen. En otras palabras, este modelo es útil porque expresa relaciones asimétricas o inversas entre Y e X porque muestra que cuando X aumenta Y disminuye hasta un punto cuyo valor es la ordenada al origen.

A.- Aplicaciones con eviews.

Ejemplos utilizando EViews 5.

La ilustración del tema de formas funcionales se realizará mediante el uso de EViews 5 utilizando datos actuales de la economía mexicana. Para ello se utilizarán las variables Consumo Privado, expresada en miles de pesos constantes de 2003, y la Tasa de Interés, tomando los CETES a 28 días, expresada en unidades porcentuales. El periodo de tiempo comprende entre el primer trimestre de 1985 hasta el cuarto trimestre de 2010.

En teoría las familias recurren al crédito para tener o incrementar su poder de compra, es decir, para consumir o incrementar su consumo. Dado que la tasa de interés es el precio que se paga por el dinero prestado, entonces a medida que ésta se incrementa se contrae la cantidad de dinero demandada en préstamo, por lo que el consumo disminuye y viceversa. Así, la relación entre el Consumo y la tasa de interés es inversa o negativa.

Ejemplo 1: Forma funcional lineal: Lin-Lin.

En este ejemplo la forma funcional que se utilizará es la lineal o Lin-Lin, donde tanto la variable dependiente (Consumo) como la independiente (CETES) se encuentran en Niveles, es decir, se trabajarán de forma tal como fueron recogidas de la fuente, sin hacérseles ninguna transformación algebraica.

Por lo que el modelo matemático planteado será el siguiente:

Consumo = f (CETES)

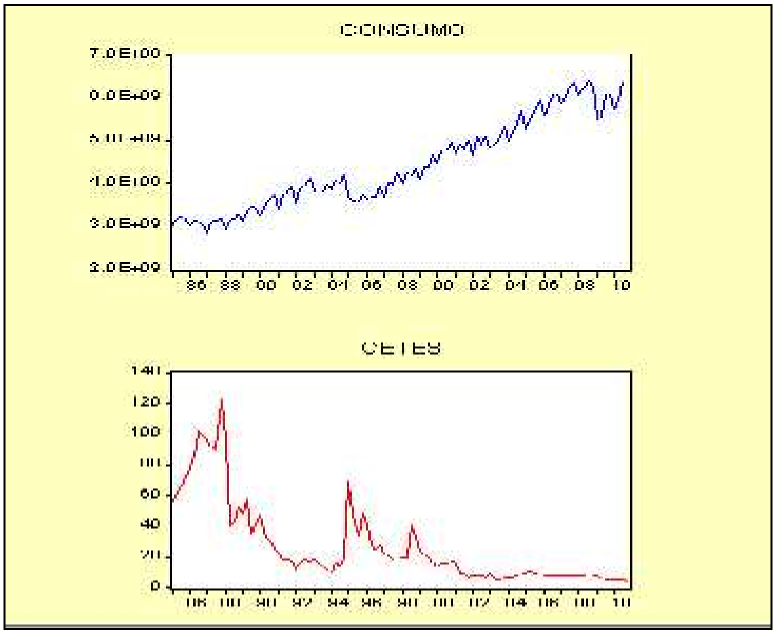
Mientras que el modelo econométrico viene dado por:

$$Con\hat{s}umo = \hat{a} - \hat{b} * CETES + \hat{e}$$

A continuación procedemos a realizar el análisis gráfico de las series (ver Cuadro 1). Por lo que respecta a la variable Consumo, observamos que ésta presenta una tendencia creciente constante, sin embargo, muestra dos grandes periodos de contracción que coinciden con la crisis económica de 1994-1995 y de 2008-2009. Por su parte, la variable CETES presenta una tendencia decreciente inconstante, asimismo, en ésta se pueden identificar claramente dos periodos: 1) El que va de 1985 a 1995 y, 2) El que va de 1995 a la actualidad. Para los dos casos de observa que la política o estrategia seguida para salir de la crisis de 1982 y 1995 fue primeramente incrementar la tasa de interés para atraer capitales al país y así estabilizar el tipo de cambio, hacer frente a los pasivos con el exterior y adquirir productos que no se elaboran internamente, para posteriormente disminuir la tasa de interés con la finalidad de canalizar dichos recursos a la esfera productiva, es decir, hacia la inversión. Llama la atención que tras la crisis económica de 2008-2009 no se haya incrementado significativamente la tasa de interés, ello en parte puede ser atribuible a la estabilidad macroeconómica del país y a las grandes reservas internacionales con las que cuenta el país, aunque sería preciso realizar un análisis más profundo sobre el particular.

Por lo que respecta al análisis estadístico de las variables (ver Cuadro 1.2) observamos que ninguna de las series se distribuye como una normal puesto que la media no es igual a la mediana. Tanto la variable Consumo como CETES no son simétricas y siguen una distribución asimétrica positiva o cargada a la derecha ya que el coeficiente de asimetría es mayor que cero. Por su parte, la variable Consumo es platocúrtica ya que el valor de la curtosis es menor a 3; y la variable CETES es leptocúrtica ya que el valor de la curtosis es mayor a 3. La no normalidad de las series se corrobora con la estadística Jarque-Bera (JB) ya que la probabilidad asociada a dicho estadístico es menor a 0.05 y su valor es igual o mayor a 6.

Gráfica 1.1 Análisis gráfico de las series.



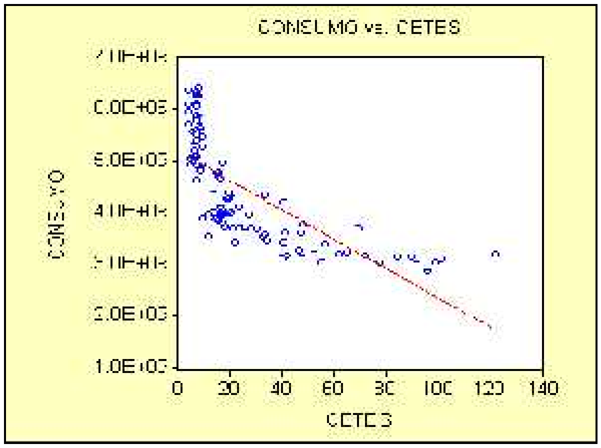
Cuadro 1.1 Análisis estadístico de las series.

	CONSUMO	CETES
Mean	4.44E+09	26.48731
Median	4.15E+09	16.55000
Maximum	6.41E+09	122.0400
Minimum	2.86E+09	4.120000
Std. Dev.	1.05E+09	26.81046
Skewness	0.580240	1.656195
Kurtosis	1.880699	4.971173
Jarque-Bera	7.935048	64.38228
Probability	0.010920	0.000000
Sum	4.62E+11	2754.680
Sum Sq. Dev.	1.14E+20	74036.46
Observations	104	104

Para determinar la relación que sigue la variable independiente con respecto a la dependiente procedemos a analizar el diagrama de dispersión (ver Cuadro 1.3). Como puede claramente apreciarse, la relación que siguen las variables no se ajusta a un comportamiento lineal, más bien éstas siguen un comportamiento polinómico o inverso, sin embargo, la relación que siguen éstas es inversa o negativa. <u>Es oportuno mencionar que el hecho de correlacionar estas dos variables mediante una relación funcional lineal se realiza con la finalidad de</u>

ilustrar la importancia de seleccionar la adecuada forma funcional que siguen las variables, interpretar los estimadores y el coeficiente de determinación.

Gráfica 1.2 Diagrama de dispersión con regresión lineal.



Ahora, procedemos a realizar la ecuación de regresión siguiendo los pasos conocidos en Eviews (ver Cuadro 1.2).

Cuadro 1.2 Resultados de la regresión

Dependent Variable: CONSUMO
Method: Least Squares
Date: 06/21/11 Time: 21:00
Sample: 1985Q1 2010Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.20E+09	39671784	52.12684	0.0000
CETES	-28569769	2651120.	-10.77649	0.0000

R-squared	0.532395	Mean dependent var	4.44E+09
Adjusted R-squared	0.527810	S.D. dependent var	1.05E+09
S.E. of regression	7.21E+08	Akaike info criterion	43.35022
Sum squared resid	5.31E+19	Schwarz criterion	43.70107
Log likelihood	-2267.811	F-statistic	116.1327
Durbin-Watson stat	0.243390	Prob(F-statistic)	0.000000

Del Cuadro 1.4 se desprende que la ecuación de regresión es:

$$Cons\hat{u}mo = 5195575277 - 28569768.73 * CETES$$

Como puede apreciarse, la relación que se obtuvo entre la Tasa de Interés (CETES) y el Consumo es inversa o negativa, tal y como se había planteado previamente en teoría. Asimismo, el valor de $\hat{a}$, que representa en este caso el Consumo autónomo, es positivo, lo cual concuerda cabalmente con la teoría keynesiana del Consumo.

En cuanto a la interpretación de los estimadores tenemos que en una forma funcional lineal éstos representan Multiplicadores o Propensiones Marginales (PMg). Así, la pendiente o el valor del estimador $\hat{b}$ <u>se interpreta como un multiplicador de la variable dependiente ante un cambio absoluto en la variable independiente</u>.

Por tanto, para nuestro ejemplo, ante un <u>incremento</u> en una unidad porcentual[13] (1%) en los CETES, el Consumo privado disminuye 28,569,768.73 miles de pesos constantes de 2003, pues tomando únicamente:

$$\hat{b} * CETES$$

Tenemos que:

$$-28569768.73 * CETES$$
$$-28569768.73 * (1)$$
$$-28569768.73$$

Por el contrario, si la Tasa de Interés <u>disminuye</u> en una unidad porcentual, entonces el Consumo crece 28,569,768.73 miles de pesos constantes de 2003, ya que:

$$-28569768.73 * CETES$$
$$-28569768.73 * (-1)$$
$$28569768.73$$

En cuanto a la ordenada al origen o Consumo Autónomo, si la Tasa de Interés fuera de cero, entonces éste ascendería a 5,195,575,277 miles de pesos constantes de 2003, puesto que:

$$Con\hat{s}umo = 5195575277 - 28569768.73 * 0$$
$$Con\hat{s}umo = 5195575277$$

[13] Aquí es importante mencionar que las unidades en las que está dada la variable independiente son porcentajes, por lo que al utilizar la forma funcional lineal el cambio que ocurre en ésta es en una unidad porcentual más no en uno por ciento. En el primer caso la variable independiente al incrementarse en una unidad porcentual pasaría de 5% a 6%, por ejemplo; y si ésta disminuyera en una unidad porcentual, entonces pasaría de 5% a 4%, por ejemplo. Mientras que en el segundo caso al incrementarse en uno por ciento la variable independiente, ésta pasaría de 5% a 5.05%, por ejemplo, ya que el 1% de 5% es 0.05%; por el contrario, si la variable independiente disminuyera en uno por ciento, entonces ésta pasaría de 5% a 4.95%, por ejemplo.

Cabe destacar que ambos estimadores obtenidos son estadísticamente diferentes de cero ya que la probabilidad asociada a ambos es menor a 0.05. Por lo que, la Tasa de Interés explica al Consumo y el valor de $\hat{a}$ es de relevancia económica.

Ahora sin usar logaritmos, procederemos a convertir el Multiplicador del Consumo ($\hat{b}$) en una elasticidad, para ello necesitamos los valores de las medias tanto de la variable dependiente como de la independiente, es decir, la media de la variable Consumo y de los CETES. Estos valores están contenidos en el Cuadro 2 (Mean). Por su parte la media del Consumo es de 4.44×10^9 y la de los Cetes es de 26.48731.

Así, sustituyendo los valores de $\hat{b}$ y de las medias de las variables en la formula siguiente, obtenemos la elasticidad del Consumo respecto a la Tasa de Interés (CETES):

$$Elasticidad = \hat{b} * \left(\frac{\bar{X}}{\bar{Y}} \right) = \hat{b} * \left(\frac{\overline{CETES}}{\overline{Consumo}} \right)$$

$$Elasticidad = -28569768.73 * \left(\frac{26.48731}{4.44x10^9} \right)$$

$$Elasticidad = -0.1704$$

La interpretación de dicho resultado es la siguiente: si la Tasa de Interés se incrementa en un 1%[14], el Consumo disminuye, en promedio, -0.1704%; por el contrario, si la Tasa de Interés disminuye en un 1%, el Consumo aumenta, en promedio, 0.1704% Por tanto, el Consumo no es muy elástico o sensible ante pequeñas variaciones en la Tasa de Interés.

En cuanto a la prueba de bondad de ajuste del modelo, el R^2 asciende a 0.5323 o 53.23% (ver Cuadro 4), lo que significa que el 53.23% de los cambios en el Consumo se explican por cambios en la Tasa de Interés (CETES), el resto es explicado por otras variables que no fueron incluidas en forma explicita en el modelo.

Como se desprende del resultado anterior, la bondad de ajuste del modelo no es alta y ello se debe a que la elección y utilización de la forma funcional lineal no es la correcta para estas variables. Veamos los resultados que se obtienen utilizando otras formas funcionales.

[14] Recuérdese que no es lo mismo que una variable que está dada en unidades porcentuales se incremente en una unidad porcentual que en un 1%, véase al respecto la nota al pie de página anterior.

Ejemplo 2: Forma funcional logarítmica-logarítmica o doble logarítmica: Log-Log.

En este ejemplo se utilizará la forma funcional Log-Log ó doble logarítmica, por lo que ambas variables, Consumo Privado y CETES, fueron transformados de sus niveles a logaritmos naturales.

El modelo matemático planteado será el siguiente:

LConsumo = f (LCETES)

Mientras que el modelo econométrico viene dado por:

$$LCon\hat{s}umo = Log(\hat{a}) - \hat{b} * LCETES + \hat{e}$$

Si llamamos:

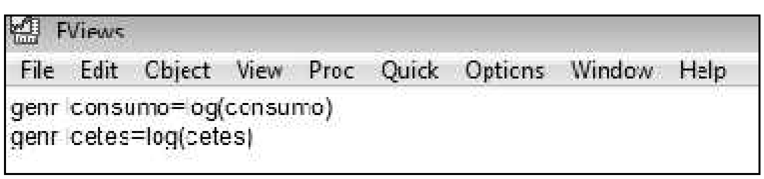
$$\hat{b}_0 \quad a \quad Log(\hat{a})$$

Entonces el modelo econométrico está dado por:

$$LCon\hat{s}umo = \hat{b}_0 - \hat{b}_1 * LCETES + \hat{e}$$

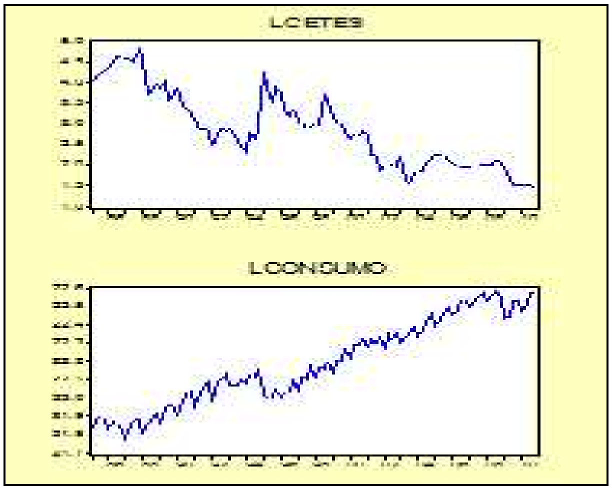
A continuación se mostrará cómo se transforman las series originales a su forma logarítmica. En la barra de comandos de EViews 5 se escribirá Genr lconsumo=Log(consumo) y oprimimos la tecla "enter" para transformar la serie del Consumo en logaritmos, con lo cual se habrá generado dicha serie, <u>la variable convertida en logaritmo aparecerá en la pantalla de Archivo de Trabajo</u>; ahora para la serie de los CETES será lo mismo, pero escribiremos: Genr lcetes=Log(cetes) y oprimimos la tecla "enter", con lo cual se habrá generado dicha serie (ver Cuadro 2.1).

Cuadro 2.1 Transformación de las series.



Ahora procedemos a realizar el análisis gráfico de las series transformadas en logaritmos (ver Cuadro 2.2). En el caso de la serie de LCETES se puede observar que tiene un comportamiento <u>decreciente no constante</u>, existen puntos mínimos en los periodos del 1991 a 1993, también en el 2001 se encuentra un punto mínimo, debido a la recesión de 2001-2002, para el 2009 hubo una disminución también pero no tan significativa, lo quiere decir que la política monetaria no fue muy activa para solucionar la crisis. Para el caso de LConsumo, ésta tiene un comportamiento creciente constante, existe un único punto mínimo para 1994 debido a la crisis de 1994, para 2009, debido a la crisis, hubo una disminución del consumo, pero no de manera muy significativa.

Gráfica 2.1 Análisis gráfico de las series.



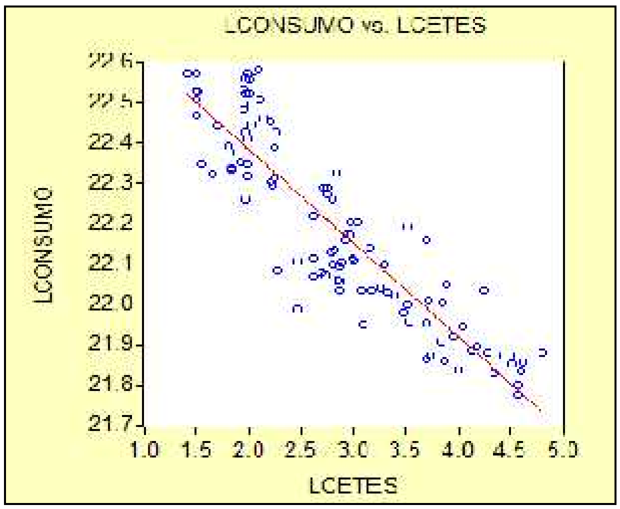
Por lo que respecta al análisis estadístico de las series encontramos que la media y la mediana no son iguales, por lo tanto las variables no se distribuyen cómo una normal. Se desprende también de dicho análisis que éstas no son simétricas y que están cargadas a la derecha ya que el coeficiente de asimetría es mayor que cero. Asimismo, ambas series son platocúrticas ya que el valor de la curtosis es menor a 3. Se corrobora la no normalidad de las series con la probabilidad del estadístico Jarque-Bera (JB) puesto que es menor a 0.05 (ver Cuadro 2.2.).

Cuadro 2.2 Análisis estadístico de las series.

	LCETES	LCONSUMO
Mean	2.849090	22.18628
Median	2.806368	22.14713
Maximum	4.804349	22.58093
Minimum	1.415853	21.77424
Std. Dev.	0.911489	0.234740
Skewness	0.387508	0.116201
Kurtosis	2.094565	1.790544
Jarque-Bera	6.155330	6.572782
Probability	0.046067	0.037389
Sum	296.3053	2307.373
Sum Sq. Dev.	85.57364	5.675603
Observations	104	104

En la siguiente gráfica 2.2 se puede ver la relación que tiene la variable dependiente con respecto a la variable independiente; se observa con el diagrama de dispersión que el comportamiento es lineal con pendiente negativa. Es importante mencionar que correlacionar estas variables mediante una relación funcional doble logarítmica se hace para ejemplificar la importancia de seleccionar la forma funcional adecuada, para interpretar los estimadores y el coeficiente de determinación.

Gráfica 2.2 Diagrama de dispersión con regresión lineal.



Procedemos ahora a correr la regresión, para ello, en la barra de comandos del programa escribimos: ls lconsumo c lcetes, y oprimimos la tecla enter; con lo cual se desplegará una ventana como la que se muestra en el Cuadro 2.3.

Cuadro 2.3 Resultados de la regresión.

Dependent Variable: LCONSUMO
Method: Least Squares
Forma funcional: log-log
Sample: 1985Q1 2010Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	22.84714	0.033129	689.6496	0.0000
LCETES	0.231954	0.011080	20.93485	0.0000

R-squared	0.811205	Mean dependent var	22.18628
Adjusted R-squared	0.809354	S.D. dependent var	0.234740
S.E. of regression	0.102495	Akaike info criterion	-1.690969
Sum squared resid	1.071526	Schwarz criterion	-1.648115
Log likelihood	90.34638	F-statistic	438.2685
Durbin-Watson stat	0.526753	Prob(F-statistic)	0.000000

La ecuación de regresión es la siguiente:

$$L\hat{Consumo} = 22.84714 - 0.231954 * LCETES$$

Como puede verse en la ecuación anterior, la relación que se obtuvo entre el logaritmo de la Tasa de Interés (LCETES) y el logaritmo del Consumo (LConsumo) es inversa o negativa, tal y como se había planteado previamente en teoría. Asimismo, el valor de $\hat{b}_0$, es la ordenada al origen que representa en este caso el Consumo autónomo, es positivo, lo cual concuerda completamente con la teoría keynesiana del Consumo.

En cuanto a la interpretación de los coeficientes obtenidos, en este caso, es decir, cuando se utiliza la forma funcional doble logarítmica, los coeficientes de los estimadores representan o miden la elasticidad de la variable dependiente con respecto a la variable independiente. En este sentido, el valor de la pendiente, $\hat{b}_1$, se interpreta como un cambio relativo, porcentual o proporcional en la variable dependiente ante un pequeño cambio, también relativo, en la variable independiente.

Por lo tanto, para nuestro ejemplo, si los CETES se incrementan en un 1%, entonces el Consumo disminuye 0.2319%; y viceversa, si la Tasa de Interés disminuye en un 1%, el Consumo se incrementa 0.2319%. Así, diremos que el Consumo es poco sensible (inelástico) a variaciones en la Tasa de Interés.

En cuanto a la interpretación de la ordenada al origen, $\hat{b}_0$, téngase presente que en este tipo de forma funcional ésta es igual a:

$$\hat{b}_0 = Log(\hat{a})$$

Para fines interpretativos lo que nos interesa es el valor de $\hat{a}$, por lo que deberemos obtener el antilogaritmo de $\hat{b}_0$. Así:

$$\hat{a} = Anti\log aritmo(\hat{b}_0)$$

En nuestro ejemplo, el valor de $\hat{b}_0 = 22.84714$, por lo que el valor de $\hat{a}$ es igual a:

$$\hat{a} = Anti\log aritmo(22.84714) = 8{,}363{,}476{,}275$$

Como se sabe, $\hat{a}$ representa el Consumo Autónomo en nuestro modelo. Por lo que, si la Tasa de Interés fuera de cero, entonces éste ascenderá a 8, 363,476,275 miles de pesos constantes de 2003.

Por otra parte, cabe destacar que ambos estimadores obtenidos son estadísticamente diferentes de cero ya que la probabilidad asociada a ambos es menor a 0.05. Por lo que, la Tasa de Interés explica al Consumo y el valor de $\hat{a}$ es de relevancia económica.

A continuación procederemos a realizar la conversión de los estimadores obtenidos. En este caso transformaremos la elasticidad a multiplicador, sin usar logaritmos. Para ello es necesario conocer los valores de la media tanto de la variable independiente como de la dependiente, es decir, la media del Consumo y de los CETES. Tales datos se encuentran contenidos en el Cuadro 1.1 (Mean). El valor de la media para el Consumo es igual a $4.44x10^9$ y para los CETES es de 26.48731.

Así, sustituyendo el valor de $\hat{b}$ y los valores de las medias de las variables en la formula siguiente, obtenemos el multiplicador del Consumo respecto a la Tasa de Interés (CETES):

$$Multiplicador = \hat{b} * \left(\frac{\bar{Y}}{\bar{X}} \right) = \hat{b} * \left(\frac{Con\bar{s}umo}{CE\bar{T}ES} \right)$$

$$Multiplicador = -0.231954 * \left(\frac{4.44x10^9}{26.48731} \right)$$

$$Multiplicador = -38{,}881{,}855.5$$

La interpretación del resultado anterior es la siguiente: si la Tasa de Interés (CETES) se incrementa en una unidad porcentual, el Consumo disminuye, en promedio, 38,881,855.5 miles de pesos constantes de 2003; por el contrario, si la Tasa de Interés disminuye en una unidad porcentual, el Consumo aumenta, en promedio, 38,881,855.5 miles de pesos constantes de 2003.

En cuanto a la bondad de ajuste, se puede ver en el Cuadro 5 que el R^2 es de 0.8112, es decir, el 81.12% de los cambios en el Consumo se explican por cambios en la Tasa de Interés (CETES), el resto es explicado por otras variables que no fueron incluidas en el modelo.

De lo anterior podemos ver que la elección de la forma funcional es adecuada, ya que explica en gran medida el comportamiento de las variables, la elección de la forma funcional log-log se ajusta a la teoría económica establecida.

Ejemplo 3: Forma funcional semilogarítmica, modelo Log-Lin.

En este ejemplo se utilizará la forma funcional logarítmica lineal (LOG-LIN) en la que la variable dependiente, Consumo, está dada en forma logarítmica (LCons); mientras que la variable independiente, CETES, está dada en forma lineal, tal y como se tomó de la fuente.

El modelo matemático se puede expresar de la siguiente forma:

$$\text{LConsumo} = f\ (CETES)$$

Mientras que el modelo econométrico se puede escribir:

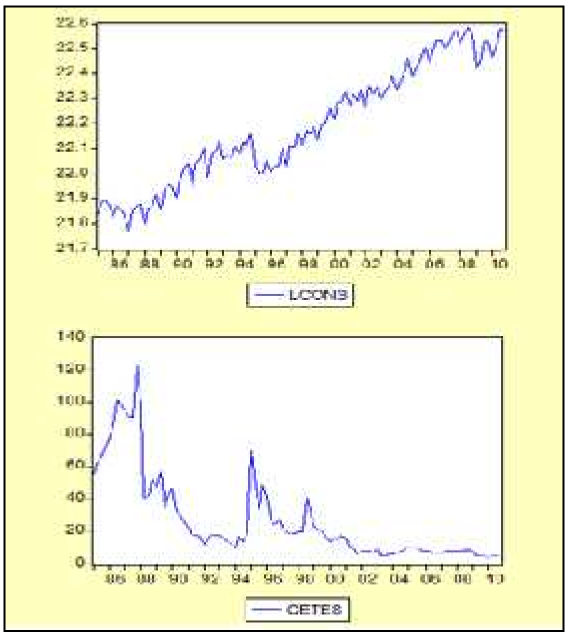
$$L\,Consumo = \hat{\alpha} - \hat{\beta} * CETES + \hat{e}$$

Como se puede observar, es necesario introducir a la variable Consumo un logaritmo, lo cual se efectúa en EViews como se explicó en el ejemplo anterior (ver forma funcional LOG-LOG).

Posteriormente se realiza el análisis gráfico de las series (ver Gráfica 3.1.). Siendo el análisis de la variable LConsumo la primera. Como se puede observar la serie presenta un comportamiento fluctuante y ascendente en lo general, teniendo una caída en 1994, año de la crisis financiera, de la cual se comenzó a recuperar para 1997, a partir de ese año la tendencia siguió siendo ascendente y creció a niveles que no había llegado antes, presentando el pico más alto en 2008, para presentar una caída en 2009 y llegar al nivel de 2004, al final de la serie, en 2010 se recupera la tendencia y se llega al nivel de 2007. La variable independiente, CETES, presenta un comportamiento inconstante y descendente en lo general. Al principio de la serie, en 1988 se observa la caída más significativa de toda la serie, la cual puede ser explicada por la poca confianza política que presentó el país cuando se cayó el sistema de conteo votos en los comicios para elección presidencial, inmediatamente después se tomaron medidas para evitar la caída de la tasa de interés, sin embargo siguió cayendo hasta encontrar su punto mínimo en 1994, año de la crisis financiera, en 1995 tuvo un repunte importante, empero, la tendencia siguió a la baja, teniendo un mínimo nuevamente en 2001, año en que estalló la crisis financiera provocada por el auge en los mercados de la informática mundial, centrada en E.U.A. En adelante la tendencia es a la baja y la crisis de 2008 no se observa con grandes afectaciones.

Por lo que respecta al análisis estadístico de las series (ver Cuadro 3.1) se puede observar que ninguna de las dos variables se comporta como una normal debido a que la media y mediana son diferentes entre si para cada caso. La dispersión o asimetría es diferente de cero para ambas variables, sin embargo la variable CETES está cargada a la derecha, más que la variable LCons, la cual se carga ligeramente a la derecha. La curtosis para la variable LCons es menor a 3, con un valor de 1.79, teniendo como resultado una distribución platocúrtica. Por otra parte, la curtosis para la variable CETES es mayor a 3, con un valor de 4.97, por lo que se obtiene una distribución leptocúrtica. La falta de normalidad en las variables se contrasta con los valores de la prueba Jarque-Bera, para la cual la regla de decisión es rechazar la presencia de normalidad si el valor es mayor que 6. Los valores bajo los que se rechaza la presencia de normalidad son, para LCons 6.572782 y CETES 64.38228, con probabilidades menores a 5%, respectivamente.

Gráfico 3.1 Análisis gráfico.

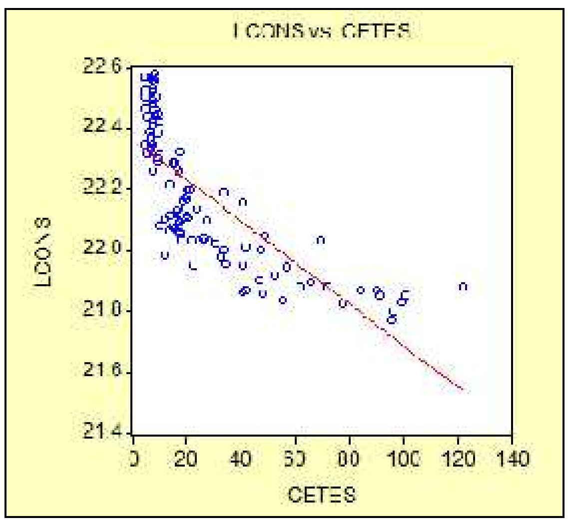


Cuadro 3.1. Análisis estadístico de las series.

	LCONS	CETES
Mean	22.18628	26.48731
Median	22.14713	16.56000
Maximum	22.58093	122.0400
Minimum	21.77424	4.120000
Std. Dev.	0.234740	25.31046
Skewness	0.116201	1.656195
Kurtosis	1.790544	4.971173
Jarque-Bera	6.572782	64.38228
Probability	0.037389	0.000000
Sum	2307.373	2754.680
Sum Sq. Dev.	5.675603	74036.46
Observations	104	104

Para determinar la relación de la variable independiente con respecto de la dependiente se procederá a realizar el análisis del diagrama de dispersión (ver Cuadro 3.2). Antes de continuar, es preciso comentar que el diagrama de dispersión siguiente es similar al diagrama de dispersión mostrado en el ejemplo de la forma funcional lineal (ver Gráfica 1.2), lo único que cambia es la escala del eje de las ordenadas, pues ahora la variable consumo se ha transformado en logaritmos para disminuir su dispersión. Así, observamos en la Gráfica 3.2 que la relación entre las variables LConsumo y CETES es inversa, tal y como se estableció en la teoría; sin embargo, su comportamiento no se ajusta a una línea recta. Como se comento anteriormente, el hecho de relacionar estas dos variables mediante una forma funcional Log-Lin se realiza con la finalidad de ilustrar la correcta especificación del modelo y la selección de la forma funcional adecuada al comportamiento de nuestras variables.

Gráfica 3.2. Diagrama de dispersión con regresión lineal.



A continuación procedemos a realizar la regresión siguiendo los pasos utilizados con el programa EViews (ver Cuadro 3.2).

Del Cuadro 3.2 se desprende la ecuación de regresión, la cual es:

$$L\widehat{Consumo} = 22.36600 - 0.006785 * CETES$$

Como se puede observar la relación que se obtuvo entre el crecimiento del consumo (LConsumo) y los CETES es negativo o inverso. El valor de $\hat{\alpha}$, el cual representa al consumo autónomo es positivo, de acuerdo con lo planteado por la teoría keynesiana del consumo. Por tanto, los resultados obtenidos cumplen con la teoría establecida. Asimismo, nótese que los estimadores son estadísticamente significativos y por ello los CETES si explican al consumo privado.

Ahora bien, los estimadores obtenidos mediante una forma funcional Log-Lin representan semielasticidades. Es decir, ante un cambio absoluto en la variable independiente ocurre un cambio relativo en la variable dependiente. Ello se debe a que ésta última se encuentra en logaritmos mientras que la primera se encuentra en niveles o valores originales. Así, para poder interpretar el valor del estimador $\hat{b}$, éste se multiplica por 100 para obtener el cambio relativo que ocurre en la variable dependiente ante el cambio absoluto que ocurre en la dependiente.

Cuadro 3.2 Resultados de la regresión.

Dependent Variable: LCONS
Method: Least Squares

Sample: 1985Q1 2010Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	22.36600	0.020600	1085.750	0.0000
CETES	-0.006785	0.000548	-12.38343	0.0000

R-squared	0.600547	Mean dependent var	22.18628
Adjusted R-squared	0.596631	S.D. dependent var	0.234740
S.E. of regression	0.149087	Akaike info criterion	-0.949535
Sum squared resid	2.267135	Schwarz criterion	-0.898682
Log likelihood	51.37584	F-statistic	153.3494
Durbin-Watson stat	0.310445	Prob(F-statistic)	0.000000

Así, para nuestro ejemplo multiplicamos:

$$\hat{b} * 100$$
$$-0.006785 * 100$$
$$-0.6785\%$$

Interrpretación: De acuerdo a Gujarati (2004: 173): "100 por $\hat{b}$ se conoce en la literatura como la semielasticidad de Y respecto a X". En efecto, dicha operación da como resultado el cambio porcentual o la tasa de crecimiento en la variable dependiente, Y, ocasionada por un cambio absoluto en la variable independiente, X. Cabe destacar que con este tipo de modelos, a decir de Wooldridge (2009: 43) "se obtiene, aproximadamente, un efecto porcentual constante".
Así, la interpretación del resultado anterior es la siguiente: si la variable independiente, en este caso CETES, se incrementa en una unidad porcentual[15],

[15] En este caso la variable independiente se incrementa en una unidad porcentual ya que está expresada originalmente en términos relativos, es decir, tal y como se tomó de la fuente. En caso de que la variable independiente estuviera dada en, por ejemplo, millones de pesos, entonces diríamos que ésta se incrementa o disminuye en un millón de pesos.

entonces el Consumo disminuye en 0.6785%, promedio trimestral; por el contrario, si la tasa de interés (CETES) disminuye en una unidad porcentual, entonces el Consumo se incrementa en 0.6785%, promedio trimestral.

Continuando con la interpretación, de acuerdo al resultado obtenido podemos afirmar que el Consumo, la variable dependiente, se incrementa a una tasa promedio trimestral de 0.6785%, en caso de que los CETES, la variable independiente, disminuya en una unidad original. Esto es igual a un crecimiento anual promedio del Consumo[16] de:

$0.6785\% * 4 = 2.714\%$, donde 4 es el número de trimestres en el año

Asimismo, en caso de que los CETES se incrementaran en una unidad porcentual, entonces la tasa de decrecimiento del Consumo trimestral será de 0.6785% y la tasa de decrecimiento anual será de:

$$-0.6785\% * 4 = -2.714\%$$

En cuanto a la interpretación del intercepto u ordenada al origen, $\hat{a}$, de acuerdo con Wooldrige (2009: 44) el valor del intercepto no tiene mucho significado ya que da el log(Y), en nuestro caso el log(Consumo), predicho, cuando X = 0, en nuestro caso cuando CETES=0. Sin embargo, de acuerdo con Gujarati (2004: 174) el valor del intercepto es importante ya que representa el valor inicial de la variable dependiente, Y, al comienzo del periodo de análisis.

Por tanto, cuando la variable independiente, X, es igual a cero, entonces:

$$\ln \hat{Y} = \hat{a}$$

Para obtener el valor de Y debemos calcular:

$$\hat{Y} = Anti\log(\hat{a})$$

En nuestro ejemplo:

$$Con\hat{s}umo = Anti\log(\hat{a})$$
$$Con\hat{s}umo = Anti\log(22.36)$$
$$Con\hat{s}umo = 5{,}138{,}361{,}031$$

La interpretación de dicho valor representa el punto inicial o de partida del Consumo al principio del periodo de análisis, es decir, es el Consumo autónomo cuando el valor que toma la variable independiente, en este caso los CETES,

[16] Se multiplica la tasa promedio trimestral por 4 ya que un año tiene cuatro trimestres.

asciende a cero. Por tanto el Consumo autónomo en nuestro ejemplo es de 5,138,361,031 miles de pesos constates de 2003 (véase el resultado obtenido para el consumo autónomo mediante el uso de la forma funcional lineal, Cuadro 1.2 del ejemplo 1).

Por otra parte, como se comento anteriormente, en la forma funcional Log-Lin $\hat{b}$ representa una semielasticidad. Para transformar dicha semielasticidad en multiplicador (pendiente) debe realizarse la siguiente transformación:

$$Pendiente = \hat{b} * \overline{Y}$$

Para nuestro caso:

$$Pendiente = \hat{b} * \overline{Consumo}$$

El valor de la media de la variable dependiente (Consumo) es de $4.44x10^9$ (ver Cuadro 1.1) y $\hat{b}$ = -0.006785 (ver Cuadro 3.2). Por tanto:

$$Pendiente = -0.006785 * 4.44x10^9$$
$$Pendiente = -30{,}125{,}400$$

Cuya interpretación es la siguiente: Si la variable independiente, CETES, se incrementa en una unidad porcentual, el Consumo disminuye en 30,125,400 miles de pesos constantes de 2003; por el contrario, si los CETES disminuyen en una unidad porcentual, entonces el Consumo se incrementa en 30,125,400 millones de pesos constantes de 2003.

Ahora, para transformar la semielasticidad en elasticidad debe realizarse la siguiente transformación:

$$Elasticidad = \hat{b} * \overline{X}$$

Para nuestro caso:

$$Elasticidad = \hat{b} * \overline{CETES}$$

El valor de la media de la variable independiente (CETES) es de 26.48731 (ver Cuadro 1.1 y 3.2) y $\hat{b}$ = -0.006785. Por tanto:

$$Elasticidad = -0.006785 * 26.48731$$
$$Elasticidad = -0.1797$$

La interpretación es la siguiente: Si la variable independiente, CETES, se incrementa en 1%, entonces la variable dependiente, el Consumo, disminuye

0.1797%; mientras que, si los CETES disminuyen en 1%, entonces el Consumo se incrementa 0.1797%. En este contexto diríamos que el Consumo no es muy sensible a las variaciones que experimenten los CETES, sin embargo, si lo afectan.

Por lo que respecta a la prueba de bondad de ajuste, dada por el coeficiente de R^2 (ver Cuadro 3.2), el cual asciende a 0.600547, que en términos porcentuales corresponde a 60.0547%, indica que los cambios en el Consumo se explican por la tasa de interés en un 60.05%, el resto se explica por otras variables que no fueron especificadas en el modelo.

Ejemplo 4: Modelo de crecimiento: Log-Lin.

Un caso particular de la forma funcional Log-Lin consiste en relacionar mediante el análisis de regresión y correlación la variable dependiente en logaritmos frente al tiempo. En este caso la variable independiente es el tiempo. A este tipo de modelos se les conoce con el nombre de "modelos de crecimiento".

En este sentido, el modelo matemático planteado posee la siguiente estructura:

$$Log(Y) = f(Tiempo)$$

Para nuestro ejemplo el modelo queda planteado de la siguiente manera:

$$LCon\hat{s}umo = f(Tiempo)$$

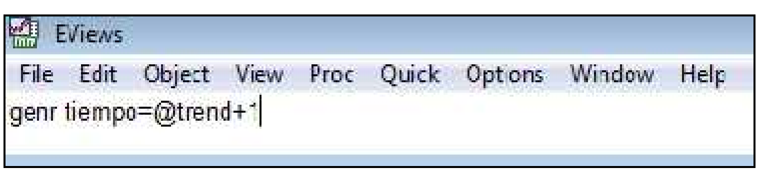
Mientras que el modelo econométrico es:

$$LCon\hat{s}umo = \hat{a} + \hat{b} * Tiempo + e$$

Nótese que en el modelo anterior el signo esperado del estimador $\hat{b}$ es positivo, ya que el resultado que se espera representa la tasa de crecimiento promedio de la variable dependiente, de ahí, precisamente, el nombre del modelo.

Ahora bien, para generar la variable tiempo, en EViews en la barra de comandos escribimos: genr tiempo=@trend+1, y oprimimos la tecla "enter" con lo cual se habrá generado dicha serie (ver Cuadro 4.1).

Cuadro 4.1 Generación de la variable tiempo.



A continuación procedemos a realizar la ecuación de regresión, para ello en la barra de comandos del programa escribimos: ls lconsumo c tiempo, seguido de la tecla enter, con lo cual se obtendrá una ventana como la contenida en el Cuadro 4.2.

La ecuación de regresión obtenida es (ver Cuadro 4.2):

$$LCon\hat{s}umo = 21.78939 + 0.007560 * Tiempo$$

En este tipo de modelos en los que la variable independiente es el tiempo, <u>$\hat{b}$ representa la tasa de crecimiento o decrecimiento promedio de la variable dependiente</u>, según sea el signo obtenido. Así, para poder interpretar el estimador debemos realizar la siguiente operación:

$\hat{b} * 100$

Por lo que para nuestro ejemplo:

$$0.007560 * 100 = 0.756\%$$

Dicho resultado lo interpretamos de la siguiente manera: La tasa de crecimiento promedio trimestral del Consumo es de 0.756%. Ello equivale a una <u>tasa de crecimiento promedio anual de:</u>

$$0.756 * 4 = 3.024\%$$

Por lo que respecta a la interpretación de la ordenada al origen, $\hat{a}$, como se recordara cuando el tiempo es igual a cero, entonces el logaritmo del consumo es:

$$LCon\hat{s}umo = 21.78939 + 0.007560 * 0$$

$$LCon\hat{s}umo = 21.78939$$

Para poder interpretar dicho resultado debemos obtener el antilogaritmo de LConsumo:

$$Con\hat{s}umo = Anti\log(21.78939)$$

$$Con\hat{s}umo = 2{,}904{,}101{,}834$$

Este valor representa el valor inicial de la variable dependiente al comienzo del periodo de análisis. Por lo que el Consumo autónomo asciende a 2,904,101,834 miles de pesos constantes de 2003 y crece a una tasa promedio trimestral de 0.756%.

Cuadro 4.2 Resultados de la regresión.

Dependent Variable: LCONSUMO
Method: Least Squares
Date: 06/24/11 Time: 23:54
Sample: 1985Q1 2010Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	21.78939	0.011046	1972.645	0.0000
TIEMPO	0.007560	0.000183	41.39087	0.0000

R-squared	0.943808	Mean dependent var	22.18628
Adjusted R-squared	0.943257	S.D. dependent var	0.234740
S.E. of regression	0.055917	Akaike info criterion	-2.910856
Sum squared resid	0.318924	Schwarz criterion	-2.860002
Log likelihood	153.3645	F-statistic	1713.204
Durbin-Watson stat	0.025163	Prob(F-statistic)	0.000000

La prueba de bondad de ajuste asciende a 0.9438, lo que significa que el 94.38% de los cambios en el Consumo se explican por el tiempo. Asimismo, los estimadores son estadísticamente significativos.

Ejemplo 5: Forma funcional semilogaritmica, modelo Lin-Log.

En este ejemplo se utilizará la forma funcional semielástica o Lin-Log, en la cual la variable dependiente (Consumo) se encuentra en niveles o valores originales, en tanto que la variable independiente (CETES) se transformará en logaritmos con la finalidad de linealizar su comportamiento. Cabe mencionar que el logaritmo a utilizar es el natural, es decir, el logaritmo en base “e”, donde “e” equivale aproximadamente a 2.7182.

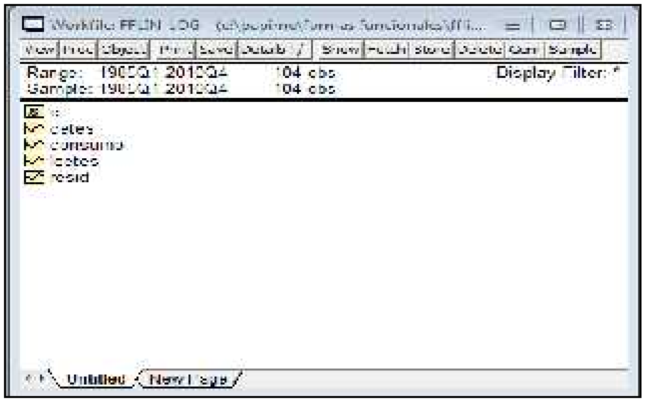
Así, el modelo matemático que se plantea es el siguiente:

Consumo = f (LCETES)

Donde “L” significa que la variable “CETES” fue convertida o transformada a logaritmos naturales.
Para poder transformar una variable en logaritmos naturales en EViews 5, debemos escribir en la barra de comandos: “genr lcetes=log(cetes)” (esto puede observase dentro del Cuadro 5.1), en seguida dar enter y la variable convertida en logaritmo aparecerá en la pantalla de Archivo de Trabajo (ver Cuadro 5.2).

Cuadro 5.1 Barra de Comandos.

Cuadro 5.2 Archivo de Trabajo (Workfile).



Por lo tanto, el modelo econométrico será el siguiente:

$$Con\hat{s}umo = \hat{a} - \hat{b} * LCETES + \hat{e}$$

Es oportuno destacar que dentro de la forma funcional Lin-Log se <u>hacen dos importantes afirmaciones económicas:</u>

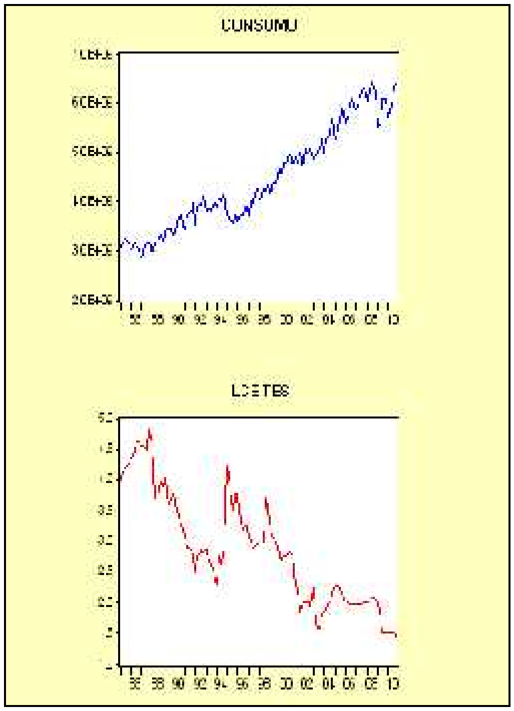
1. Existen relaciones no lineales entre la variable dependiente y la variable independiente.

2. Un incremento porcentual en la variable independiente, dará como resultado un cambio absoluto en la variable dependiente.

Por otra parte, por lo que respecta <u>al análisis gráfico</u> de las series (Gráfica 5.1), se observa que éstas tienen el mismo comportamiento, en cuanto a sus tendencias, como en el ejemplo 1 de formas funcionales, es decir, como en la forma funcional lineal. El único cambio que sobresale se da en la modificación de escala en la serie de la Tasa de Interés (CETES), ya que se procedió a transformar la variable en logaritmos (ver Cuadro 5.4).

En cuanto al análisis estadístico de las variables (Cuadro 5.3), <u>se observa que ninguna de las series se distribuye como una normal dado que la media no es igual a la mediana</u>. Tanto la variable Consumo como LCETES no son simétricas y siguen una distribución asimétrica positiva o cargada a la derecha ya que el coeficiente de asimetría es mayor que cero. Ambas variables, Consumo y LCETES, son platocúrticas debido a que el valor de la curtosis es menor a 3. La no normalidad de las series como se ha venido manejado en los ejemplos anteriores de regresión simple, múltiple y de formas funcionales, se corrobora con la estadística Jarque-Bera (JB) ya que la probabilidad asociada a dicho estadístico es menor a 0.05 y su valor es igual o mayor a 6 (ver Cuadro 5.3).

Gráfica 5.1.Análisis gráfico de las series.

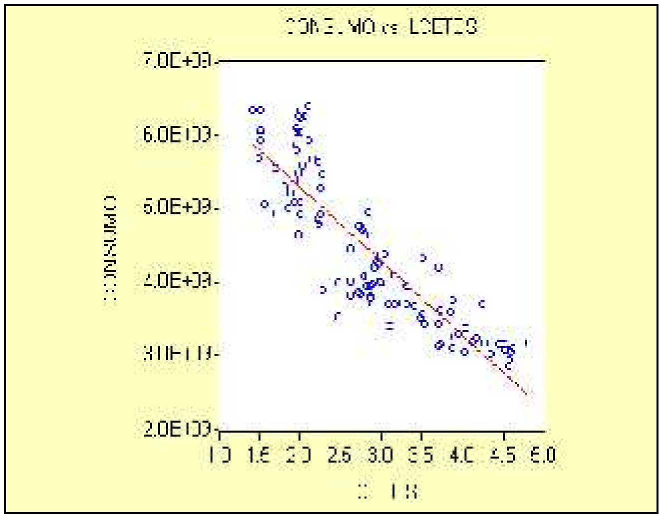


Cuadro 5.3 Análisis estadístico de las series.

	CONSUMO	LCETES
Mean	4.44E+09	2.849090
Median	4.15E+09	2.806368
Maximum	6.41E+09	4.804349
Minimum	2.85E+09	1.415853
Std. Dev.	1.05E+09	0.911489
Skewness	0.380240	0.387608
Kurtosis	1.880699	2.094565
Jarque-Bera	7.935040	6.155030
Probability	0.018920	0.046067
Sum	4.62E+11	296.3053
Sum Sq. Dev.	1.14E+20	85.57364
Observations	104	104

Ahora procedamos a determinar la relación que siguen las variables mediante el análisis al diagrama de dispersión. De la Gráfica 5.2 se desprende que la relación entre las variables no es totalmente lineal, manteniéndose un comportamiento inverso o negativo, aunque a diferencia de la forma funcional lineal, la variable independiente al estar ahora en logaritmos se ajusta mejor a dicho comportamiento lineal con respecto a la variable dependiente.

Gráfica 5.2. Diagrama de dispersión con regresión lineal.



Para la obtención de la ecuación de regresión se utilizará la variable LCETES, la cual se generó al inicio de este ejercicio; dentro de la barra de comandos se escribirá "ls consumo c lcetes" y se tendrá como resultado el siguiente:

Cuadro 5.4 Resultados de la regresión.

Dependent Variable: CONSUMO
Method: Least Squares
Date: 06/21/11 Time: 18:44
Sample: 1985Q1 2010Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.32E+09	1.64E+08	44.74002	0.0000
LCETES	-1.01E+09	96494334	-10.47276	0.0000

R-squared	0.769978	Mean dependent var	4.44E+09
Adjusted R-squared	0.767522	S.D. dependent var	1.05E+09
S.E. of regression	5.06E+08	Akaike info criterion	42.94120
Sum squared resid	2.61E+19	Schwarz criterion	42.99206
Log likelihood	-2230.943	F-statistic	341.2430
Durbin-Watson stat	0.422026	Prob(F-statistic)	0.000000

Del Cuadro 4.6 se desprende la siguiente ecuación de regresión:

$$\widehat{Consumo} = 7317958821 - 1010540198 * LCETES$$

Con la ecuación de regresión que se obtuvo puede observarse que se sigue comprobando la teoría que se planteó desde un principio, esto quiere decir que se tiene una relación inversa entre el logaritmo de la Tasa de Interés (LCETES) y el Consumo. Por lo que respecta al Consumo autónomo, $\hat{a}$, continua teniendo un comportamiento positivo, lo cual sigue siendo válido para la teoría keynesiana del Consumo. Asimismo, se destaca el hecho de que los estimadores obtenidos son estadísticamente significativos o diferentes de cero, puesto que la probabilidad asociada a ambos es menor a 0.05. Por tanto, el logaritmo de la Tasa de Interés explica al Consumo y el valor de $\hat{a}$ es de relevancia económica.

En cuanto a la interpretación de los estimadores, en una forma funcional Lin-Log éstos representan semielasticidades. Por lo que los estimadores no pueden ser interpretados de forma directa. Para poder ser interpretados éstos deben multiplicarse por 0.01, es decir, por 1% o también se pueden dividir entre 100.

Una semielasticidad en una forma funcional Lin-Log se interpreta como: ante un pequeño cambio relativo, o pequeña variación, en la variable independiente, en este caso la Tasa de Interés en logaritmos (LCETES), ocurre un cambio absoluto en la variable dependiente, en este caso el Consumo.

Así, para interpretar al estimador de la variable independiente, $\hat{b}$ se multiplica por 0.01, es decir, por 1% o también se puede dividir entre 100. Para nuestro caso, el resultado es:

$$\Delta Y = \hat{b} * (\Delta X / X) = -1010540198 \quad * (0.01) \; = -10105402$$

Dicho resultado se interpreta de la forma siguiente: Ante un incremento en un 1% en CETES, el Consumo disminuye, dado su signo negativo, en 10,105,402 miles de pesos constantes de 2003. Por el contrario, si CETES disminuye en un 1%, el Consumo se incrementará en 10,105,402 miles de pesos constantes de 2003.

Ahora bien, en cuanto al Consumo autónomo, $\hat{a}$, si el logaritmo de la Tasa de Interés (CETES), o la pequeña variación que experimentaran los CETES, fuera de cero, éste ascendería a:

$$Cons\hat{u}mo = \frac{7317958821}{100} - 10105402 * 0$$
$$Cons\hat{u}mo = 73179588$$

Es decir, el Consumo independientemente de las variaciones y/o cambios que experimente la Tasa de Interés es de 73,179,588 miles de pesos constantes de 2003.

Por otra parte, como se ha comentado, los estimadores obtenidos mediante el uso de la forma funcional Lin-Log representan semielasticidades. Sin embargo, éstos pueden ser transformados en multiplicadores o elasticidades.

Para la obtención del Multiplicador o Propensión Marginal (PMg) se hará uso de una fórmula, para ello se necesita el valor del estimador $\hat{b}$ y la media de la variable independiente, es decir, la media de CETES. El valor de la media de CETES es de 26.48731 (ver Cuadro 3.1). Así:

$$PMg = \hat{b} * \left(\frac{1}{\bar{X}} \right) = \hat{b} * \left(\frac{1}{\overline{CETES}} \right)$$

$$PMg = -1010540198 * \left(\frac{1}{26.48731} \right) = -38{,}151{,}862.08$$

Por lo tanto, si se da un incremento en una unidad porcentual en la Tasa de Interés (CETES), el Consumo privado disminuye en promedio 38,151,862.08 miles de pesos constantes de 2003; y si la Tasa de Interés cae en una unidad porcentual, el Consumo privado aumentará en 38,151,862.08 miles de pesos constantes de 2003[17].

A continuación se convertirá el estimador $\hat{b}$ en una elasticidad mediante una fórmula, en la que nuevamente se necesita el valor de la media, pero en este caso de la variable dependiente, es decir, el Consumo. El valor de la media del Consumo equivale a $4.44x10^9$. Así:

$$Elasticidad = \hat{b} * \left(\frac{1}{\bar{Y}} \right) = \hat{b} * \left(\frac{1}{\overline{Consumo}} \right)$$

$$Elasticidad = -1010540198 * \left(\frac{1}{4.44x10^9} \right)$$

$$Elasticidad = -0.2275$$

Por lo tanto, si la Tasa de Interés (CETES) se incrementa en un 1%, el Consumo disminuye, en promedio, 0.2275%; por el contrario, si la Tasa de Interés disminuye en un 1%, el Consumo aumenta en 0.2275%[18].

[17] Nótese que en la interpretación del ahora multiplicador o Propensión Marginal (PMg), la variable independiente (CETES) aumenta o disminuye en una unidad porcentual y no en un 1%. Por lo que dicha variable vuelve a estar en niveles y no en logaritmos, de ahí que se hable de CETES y no de LCETES. Asimismo, el cambio ocasionado por dicho incremento o decremento en la variable independiente, repercute en términos absolutos en la variable dependiente.

[18] Al convertir el estimador obtenido mediante el uso de la forma funcional Lin-Log, que es una semielasticidad, en elasticidad, ahora la variable independiente se incrementa o disminuye en un 1% y no en una unidad porcentual. De ahí que ahora se hable de LCETES y no de CETES. Asimismo, la variación ocasionada por dicho incremento o decremento en la variable independiente, repercute en términos relativos en la variable dependiente.

En cuanto a la prueba de bondad de ajuste tenemos que: R^2 = 0.769878 = 76.9878%, lo que significa que el 76.9878% de los cambios en el Consumo se explican por variaciones en la Tasa de Interés (LCETES), el resto es explicado por otras variables que no fueron incluidas en el modelo. Si se compara con el primer ejemplo de formas funcionales, se puede observar que hay un incremento en la bondad de ajuste, esto quiere decir, que las variables se explican de mejor forma. ***Sin embargo, si se comparan*** *las cinco formas funcionales, se notará que la Log Lin es la que produjo el mejor ajuste de los valores estimados a los valores reales u observados de la variable dependiente, en virtud de que su Coeficiente de Determinación es de 0.9438, es decir, el más alto, en otras palabras, se dice que la forma funcional Lin Lin es la que produce la menor bondad de ajuste porque su Coeficiente de determinación es es el menor: 0.53 (Cuadro 1.2).*

Ejemplo 6: Forma funcional reciproca o inversa.

En este ejercicio se trabajará el modelo reciproco, el cual tiene la siguiente forma:

$$Y_i = \hat{a} + \hat{b}\left(\frac{1}{X_i}\right) + u_i$$

Para nuestro caso la variable dependiente es el Consumo expresada en miles de pesos constantes de 2003, y nuestra variable independiente es la Tasa de Interés tomando los CETES a 28 días, expresada en unidades porcentuales. De este modo nuestra ecuación de regresión es la siguiente:

$$Consumo = \hat{a} + \hat{b}\left(\frac{1}{\text{CETES}}\right) + u_i$$

La variable CETES no es lineal en el modelo, i.e., en la variable independiente, ya que entra inversamente $\left(\frac{1}{CETES}\right)$, a pesar de ello, nuestro modelo sigue siendo lineal en los parámetros, â y $\hat{b}$.

Por otra parte, se observa que a medida de que la variable independiente, en este caso CETES, crece infinitamente, el término $\hat{b}\left(\frac{1}{CETES}\right)$ se acerca a cero y por consiguiente al ser este cada vez más cercano a cero no tiene influencia en la regresión, por lo que la variable dependiente, en este caso el Consumo, será igual al estimador â.

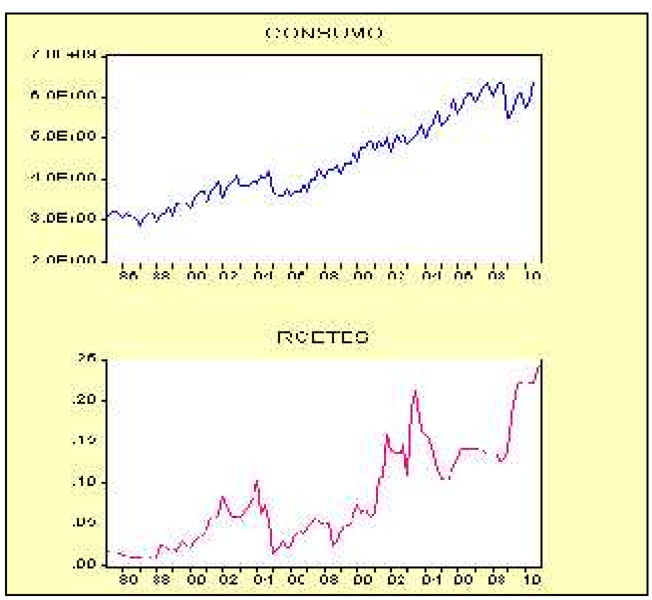
Ahora bien, la variable CETES en el modelo reciproco no es ingresada a la regresión como tal, sino que es la división de 1/CETES, a continuación se explica cómo debe ser generada en EViews.

Esta variable se obtiene de manera simple con un comando; en la barra de comandos se escribe “genr” seguida de un espacio y después el nombre que se le dará a esta variable “rcetes”, después de esto sin espacios se escribe “=” seguido de la operación a realizar, en este caso 1/cetes, cabe recalcar que la variable CETES ya fue ingresada en EViews previamente, ésta debe existir ya en el programa sino no tiene relevancia la operación y no generará ningún resultado; después se presiona Enter. El comando final es este: genr rcetes=1/cetes, (ver Cuadro 6.1).

Cuadro 6.1 Área de comandos

Después de generar el inverso o reciproco de la variable CETES proseguimos a analizar la tendencia de nuestras variables, en este caso el Consumo y RCETES como se muestra en la gráfica 6.1. Podemos observar en ella que la tendencia del Consumo es creciente con dos caídas notables que coinciden con los periodos de crisis de 1994 y la crisis más reciente del año 2008, en el primer ejemplo de formas funcionales, el modelo Lin-Lin, se ha explicado detalladamente la tendencia de la variable Consumo; para la variable “rcetes” antes generada, se observa que cambia completamente su tendencia, ya que al ser dividida se muestra inversamente y ahora la tendencia es positiva, pero es muy inestable, por ello ahora comparte con la variable dependiente la tendencia que se verá reflejada en la regresión.

Gráfica 6.1. Análisis gráfico de las series.



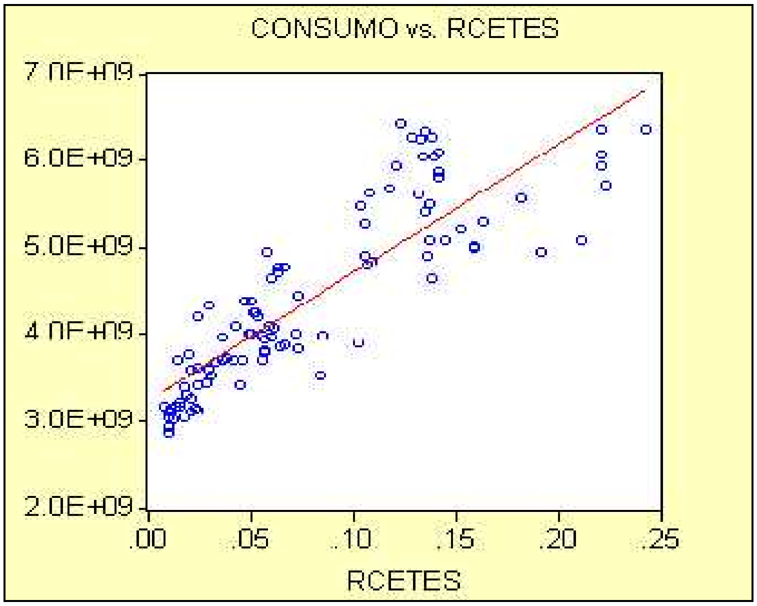
Cuadro 6.2 Análisis estadístico de las series.

	CONSUMO	RCETES
Mean	4.44E+09	0.082031
Median	4.15E+09	0.060425
Maximum	6.41E+09	0.242718
Minimum	2.86E+09	0.008194
Std. Dev.	1.05E+09	0.061924
Skewness	0.380240	0.771519
Kurtosis	1.880699	2.651547
Jarque-Bera	7.935048	10.84368
Probability	0.018920	0.004419
Sum	4.62E+11	8.531249
Sum Sq. Dev.	1.14E+20	0.394957
Observations	104	104

Pasando al análisis estadístico de las variables que se muestra en el Cuadro 6.2 observamos que ninguna variable se distribuye como una normal puesto que la media no es igual a la mediana. Asimismo, ambas variables siguen una distribución asimétrica positiva o cargada a la derecha ya que el coeficiente de asimetría es mayor que cero. Evaluando el estadístico curtosis tenemos que la variable Consumo y RCETES son platocúrticas ya que su valor es menor a 3, a diferencia del modelo Lin-Lin, tenemos que la variable RCETES cambia de ser leptocúrtica a platocúrtica por ser invertida. La normalidad de las series se puede analizar con la estadística Jarque-Bera (JB) donde concluimos que no se comportan las series como una normal ya que la probabilidad asociada a dicho estadístico es menor a 0.05 y su valor es mayor a 6.

Pasemos ahora a analizar la Gráfica 6.2 que mediante un diagrama de dispersión nos permite determinar la relación que sigue la variable RCETES con respecto a la variable Consumo. A diferencia de la regresión Lin-Lin en donde se especificaba, dado el diagrama de dispersión, que la regresión lineal no era la adecuada para el modelo ya que los datos no se ajustan a la regresión lineal, mostrando que no es el mejor método, en el caso del modelo reciproco al tener la inversa de los CETES podemos observar en el diagrama de dispersión que aunque la relación negativa entre las variables cambia, porque ahora tenemos una relación directa, la regresión tiene un mayor ajuste a los datos, aunque no perfecta, al estar los datos muy dispersos no se puede concluir que esta forma funcional es la adecuada.

Gráfica 6.2 Diagrama de dispersión con regresión lineal.



Para explicar mejor la relación que tiene la variable independiente RCETES con la variable dependiente Consumo, pasáremos al Cuadro 6.3 que muestra los resultados de la regresión.

Del Cuadro 6.3 se obtiene la siguiente ecuación:

$$Consumo = 3{,}227{,}465{,}254 + 14{,}767{,}225{,}170 * RCETES$$

Sin embargo, como se trata de un modelo reciproco, el signo obtenido para $\hat{b}$ es el inverso u opuesto, así:

$$Con\hat{s}umo = 3{,}227{,}465{,}254 - 14{,}767{,}225{,}170 * RCETES$$

Por lo tanto tenemos que la relación entre la tasa de interés (RCETES) y el Consumo es negativa, lo cual concuerda con la teoría keynesiana del consumo. Asimismo, el valor de $\hat{a}$, que representa en este caso el Consumo autónomo, es positivo, lo cual concuerda con la teoría establecida.

Cuadro 6.3. Resultados de la regresión

Dependent Variable: CONSUMO
Method: Least Squares
Sample: 1985Q1 2010Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.23E+09	84582787	38.15747	0.0000
RCETES	1.48E+10	8.24E+08	17.91236	0.0000

R-squared	0.758791	Mean dependent var	4.44E+09
Adjusted R-squared	0.756427	S.D. dependent var	1.05E+09
S.E. of regression	5.18E+08	Akaike info criterion	42.98826
Sum squared resid	2.74E+19	Schwarz criterion	43.03911
Log likelihood	-2233.389	F-statistic	320.8705
Durbin-Watson stat	0.391633	Prob(F-statistic)	0.000000

Por otra parte, como se sabe, en este tipo de modelos a medida que la variable dependiente, en este caso CETES, aumenta indefinidamente el término $\hat{b}*(1/CETES)$ se acerca a cero y la variable dependiente, en este caso, el Consumo, se aproxima a su valor asintótico, $\hat{a}$, que para nuestro caso equivale a 3,227,465,254 miles de pesos constantes de 2003. Asimismo, si la variable independiente disminuye constantemente sin llegar a tocar el cero, el término $\hat{b}*(1/CETES)$ crece y por tanto el valor de la variable dependiente disminuye, dado el signo negativo del término anterior, para este caso. Sin embargo, para fines de interpretación es necesario realizar algunas transformaciones.

Así, para obtener Multiplicadores o Propensiones Marginales (PMg), en el caso de la forma funcional reciproca los estimadores por sí mismos no ofrecen la pendiente de la regresión, por lo que debemos de realizar la siguiente transformación:

$$Pendiente = -\hat{b}*\left(\frac{1}{\overline{X}^{2}}\right)$$

Donde:

$$\overline{x} = \quad Media \quad de \quad la \quad variable \quad independiente$$

$$Pendiente = -\hat{b} * \left(\frac{1}{\overline{CETES}^2} \right) = -14{,}767{,}225{,}170 * \left(\frac{1}{26.49^2} \right)$$

$$Pendiente = \text{-}21{,}044{,}324.04$$

OJO: la media de los CETES en el Cuadro 6.2 es 0.082031, corregir arriba y abajo

Por tanto, si la tasa de interés, CETES, se incrementa en una unidad porcentual, entonces el Consumo disminuye 21,044,324.04 miles de pesos constantes de 2003; por el contrario, si la tasa de interés disminuye en una unidad porcentual, el Consumo se incrementará en 21,044,324.04 miles de pesos constantes de 2003. Lo cual concuerda completamente con la teoría del consumo, puesto que si disminuye la tasa de interés entonces las personas se ven motivadas a pedir dinero prestado e incrementar por esta vía su consumo y viceversa.

Ahora, <u>para convertir el estimador $\hat{b}$ en una elasticidad</u> se utilizan las medias de las variables Consumo y CETES, que están en el Cuadro 6.2. Por su parte la media del Consumo es de $4.44x10^9$ y la de CETES es de 0.082031. Así, sustituyendo los valores de $\hat{b}$ y de las medias de las variables en la formula siguiente, obtenemos la elasticidad del Consumo respecto a la tasa de interés:

$$Elasticidad = -\hat{b} * \left(\frac{1}{\overline{X} * \overline{Y}} \right) = -\hat{b} * \left(\frac{1}{\overline{CETES} * \overline{Consumo}} \right)$$

$$Elasticidad = -14{,}767{,}225{,}170 * \left(\frac{1}{4.44x10^9 * 0.082031} \right)$$

$$Elasticidad = \text{-}0.1255$$

Por lo que la interpretación de dicho resultado es que si la Tasa de Interés se incrementa en un 1%, el Consumo disminuye, en promedio, -0.1255%; por el contrario, si la Tasa de Interés disminuye en un 1%, el Consumo aumenta, en promedio, 0.1255%. Por lo que concluimos que el Consumo no es sensible ante pequeñas variaciones en la Tasa de Interés.

En cuanto a la prueba de bondad de ajuste tenemos que R^2 = 0.758791, (ver Cuadro 6.2) lo que significa que el 75.8791% de las variaciones del Consumo son explicadas por la variable independiente RCETES. Esta cifra es más alta que la del modelo Lin-Lin que arrojaba un R^2 = 0.5323, esto quiere decir que en el modelo reciproco existe una relación más fuerte entre las variables que en el primero, <u>*pero, como se anotó en párrafos arriba, el mejor ajuste lo produjo la forma funcional Log-Lin.*</u>

IV.14.1.- Reactivos para reafirmar los conocimientos.

A.- Marco teórico de la transformación de formas funcionales no lineales en lineales

A.-Conteste con una "X" en **SI** cuando la frase sea verdadera y también con una "X" en NO cuando la frase sea falsa:

1.-La elasticidad y R^2 no ayudan a seleccionar la forma funcional adecuada para transformar la no lineal: Si___; No____.

2.- El especialista Dominick salvadores dice que "Cuando ninguna teoría de la dispersión de puntos es de ayuda, la función lineal se trata usualmente primero debido a su simplicidad": SI_____;NO_____.

3.-Las transformaciones de funciones no lineales a lineales mas útiles y comunes son la log-log,log-lin,lin-log, reciproca y polinomial: Si_____; No________

4.- Al usar la log-log el coeficiente de la variable independiente es un estimador insesgado : SI:_______; NO:___________

5.-Al usar la log-log la elasticidad de la variable regresada se obtiene directamente de los valores de las pendientes de las regresoras: SI:__; NO__.

6.-Al usar la log-log con las pendientes de las variables regresoras no siempre se obtienen economías de escala: SI___; NO____.

7.-Al usar la log-log si las variables regresoras son factores de la producción, con sus pendientes se pueden obtener economías de escala cuando la su suma es mayor que 1: SI____; NO______:

8.-Con la forma funcional semilogarítmica log lin se puede calcular la tasa de crecimiento de la variable regresada con la pendiente de la regresora, siempre y cuando ésta represente al tiempo: SI____;NO_____.

9.- Con la forma funcional recíproca no se puede calcular la curva de Phillips: SI_____; NO____.

10.- La siguiente es la fórmula de la elasticidad del modelo log recíproco: $_2\left(\frac{1}{X}\right)$

: SI____; NO______.

V. Modelo lineal general o multiple, MLG.

Muchas veces necesitamos explicarnos los cambios de Y en función de varias variables. En este caso se aplican el análisis de regresión y correlación múltiple, que, como se viene señalando, es un método con el cual se puede hacer análisis multivariable, llamado así porque establece una relación funcional entre una variable dependiente (Y) y un grupo de variables independientes o explicativas (X,Z,Q), cuyos coeficientes de su ecuación de regresión, también se obtienen con MCO y como en el MLS, determinan el impacto que tienen las variables independientes en la dependiente.

V.1. MLS versus MLG

Como se indicó previamente, ciertos estudiosos del tema suelen decir que el MLG representa mejor la realidad económica que se desea estudiar econométricamente, lo cual es cierto porque participan más variables explicativas. Con ese enfoque el MLS se reduciría a simple instrumento didáctico para introducir la explicación de la teoría econométrica. Creo que eso no es del todo exacto, dado que finalmente todo depende de la manera en que se formule la teoría económica, ya que si por ejemplo el investigador establece que la tributación fiscal depende de los ingresos de las personas físicas y morales de México, pienso que aquí aplica muy bien el MLS.

Por el contrario el MLG es conveniente si la teoría económica se plantea como el siguiente ejemplo: Las ventas Y de los ocho vendedores del MLS, ahora decimos que pueden explicarse no solamente en función de sus años de experiencia sino también digamos, de las calificaciones que obtienen los vendedores en sus cursos de actualización quedando la función $\boldsymbol{Y=f(X_1, X_2)}$.

Así, si determinamos que $\boldsymbol{Y=f(X_1, X_2)}$. Donde Y=ventas, X_1=años de experiencias; X_2= calificaciones de los vendedores, en un momento dado **(datos de corte transversal).**

En este caso la ecuación de regresión múltiple en la población será: $\boldsymbol{Y=\alpha+\beta X_1+\pi X_2}$, cuyos valores de sus parámetros α, β y π se hallan con sus estimadores correspondientes de la muestra a, b, c[19]. Al ser tres, para encontrar sus valores es necesario resolver el siguiente sistema de ecuaciones:

$$\sum Y = na + b\sum X_1 + c\sum X_2 \qquad (1)$$

$$\sum YX_1 = a\sum X_1 + b\sum X_1^2 + c\sum X_1X_2 \qquad (2)$$

$$\sum YX_2 = a\sum X_2 + b\sum X_1X_2 + c\sum X_2^2 \qquad (3)$$

[19] Recuérdese que los estimadores α y β (alfabeto griego) representan el universo, mientras que â, ƃ y ĉ (abecedario) representan a la muestra. En este caso α ≠ al nivel de significación.

Para ilustrar el uso de tres variables en el caso del coeficiente de correlación múltiple, podemos introducir el subíndice $Y=Y_1$, tal que se representa con r_{123} y generalizando hacia el *coeficiente de determinación, ahora* lo denominamos por: R^2_{123}: e indica que los cambios en Y_1, se deben a las variaciones de X_1 y X_2.

V.2. Ejemplo de cómo se maneja con datos de series de tiempo.

Con el fin de ilustrar el uso de este instrumental en el desarrollo de una teoría económica a continuación se presenta el siguiente caso:

Con base en este enfoque a continuación se expone un ejercicio que permitirá ilustrar: Como se prueba la teoría económica de que el consumo (Y) depende del ingreso disponible (después de impuestos) (X_1) y de la inflación (X_2); de X_1 en forma positiva y de X_2 en forma negativa.

V.2.1 Obtención de la ecuación de regresión.

Si decimos que $Y=\alpha+\beta X_1+\pi X_2+U_i$ es la ecuación de la regresión, entonces para desarrollar el modelo diremos que los datos en los últimos 10 años están expresados en millones de pesos para Y, X_1, y que están en porcentajes para X_2. De esta manera la ecuación de regresión que servirá para probar con la muestra la teoría económica quedará planteada de la siguiente manera:

$$\hat{y} = \hat{a} + \hat{b}X_i + \hat{c}X_2 + e_i$$

Dónde:

$\hat{y}$ = Y calculada o estimada

e_i **expresa que no hay una relación exacta** de Y con respecto a "X_1 "y" X"$_2$ por lo que
e_i es conocida como el complemento residual.

Así, para encontrar $\hat{y}$ se requiere conocer los valores de $\hat{a},\hat{b},\hat{c}$ mismos que se determinan con las tres ecuaciones normales siguientes:

$$\sum \hat{Y} = n\hat{a} + \hat{b}\sum X_1 + \hat{c}\sum X_2.$$
$$\sum \hat{Y}X_1 = \hat{a}\sum X_1 + \hat{b}\sum X_1^2 + \hat{c}\sum X_1X_2$$
$$\sum \hat{Y}X_2 = \hat{a}\sum X_2 + \hat{b}\sum X_1X_2 + \hat{c}\sum X_2^2$$

Cuando se expresan en forma de desviaciones con respecto a la media aritmética se pueden resolver simultáneamente para $\hat{b}$ y $\hat{c}$, así:

$$\hat{b} = \frac{(\sum x_1 y)(\sum x_2^2) - (\sum x_2 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\hat{c} = \frac{(\sum x_2 y)(\sum x_1^2) - (\sum x_1 y)(\sum x_1 x_2)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

tal que:

$$\hat{a} = \overline{Y} - \hat{b}\overline{X}_1 - \hat{c}\overline{X}_2$$

Se dice que $\hat{b}$ *y* $\hat{c}$ son estimadores óptimos lineales insesgados, es decir:

$\in(\hat{b})$= β parámetro de la población
$\in(\hat{c})$ = π parámetro de la población
Recuérdese que $\in$ es el símbolo de la esperanza matemática.
En esta regresión múltiple $\hat{b},\hat{c}$ son estimadores parciales de β y π.

Una vez establecidos tanto la teoría económica como la metodología para probarla, MCO, con los datos que aparecen en la tabla 1, hacemos los siguientes cálculos para obtener $\hat{a},\hat{b},\hat{c}$.

Tabla 1: Relación del consumo (Y) con el ingreso (X_1) y la inflación (X_2) (del año 1 al 10)

Años	**Y**	**X_1**	**X_2**	***Y***	***x_1***	***x_2***	***yx_1***	x_1^2	x_2^2	**$x_1 x_2$**	**$x_2 y$**
1	3	1	8	-3.1	-4.65	+0.9	14.415	21.623	0.81	-4.185	-2.79
2	2	2	15	-4.1	-3.65	+7.9	14.965	13.323	62.41	-28.835	-32.39
3	4	2.5	10	-2.1	-3.15	+2.9	6.615	9.923	8.41	-9.135	-6.09
4	5	3	9	-1.1	-2.65	+1.9	2.915	7.023	3.61	-5.035	-2.09
5	5	4	7	-1.1	-1.65	-0.1	1.815	2.723	0.01	+0.165	+0.11
6	7	5	6	+0.9	-0.65	-1.1	-0.585	0.423	1.21	+0.715	-0.99
7	6	7	8	-0.1	+1.35	+0.9	-0.135	1.823	0.81	+1.215	-0.09
8	8	8	4	+1.9	+2.35	-3.1	4.465	5.523	9.61	-7.285	-5.89
9	9	9	3	+2.9	+3.35	-4.1	9.715	11.223	16.81	-13.735	-11.89
10	12	15	1	+5.9	+9.35	-6.1	55.165	87.423	37.21	-57.035	-35.99
n=10	**61**	**56.5**	**71**	***0***	***0***	***0***	***109.349***	***161.03***	***140.9***	***-123.15***	***-98.10***

Nota: Y, X_1, X_2 son los valores originales, en tanto que y, x_1 y x_2 son minúsculas e indican las desviaciones de los términos con respecto a sus medias aritméticas respectivas.

Cuyos valores de sus medias son:

$\overline{Y} = 6.1$

$\overline{X}_1 = 5.65$

$\overline{X}_2 = 7.1$

Sustituyendo del Cuadro anterior, tenemos:

$$\hat{b} = \frac{(109.349)(140.09) - (-98.10)(-123.15)}{(161.03)(140.09) - (-123.15)^2} = \frac{15.318.7 - 12,081.01}{22,558.69 - 15,165.93} = \frac{3,237.69}{7,392.77} = 0.4379$$

$$\hat{c} = \frac{(-98.1)(161.03) - (109.349)(-123.15)}{(161.03)(140.09) - (123.15)^2} = \frac{-15,797.04 + 13,466.33}{22,558.69 - 15,165.92} = \frac{-2,330.71}{7,392.77} = -0.3152$$

$$\hat{a} = 6.1 - (0.4379)(5.65) - (-0.3152)(7.1) = 6.1 - 2.4741 + 2.2379 = 5.8638$$

Observamos que $\hat{b}$= 0.4379 tiene signo positivo y $\hat{c}$= -0.3152 tiene signo negativo. Estos resultados confirman la teoría económica, motivo por el cual podemos continuar probando la relación que tiene el consumo (Y) con el ingreso (X_1) y la inflación (X_2). Por consiguiente, la ecuación de regresión múltiple es:

$\hat{y}$ *= 5.8638 + 0.4379*X_1*- 0.3152*X_2

V.2.2.- Obtención de los valores residuales.

Ahora bien en virtud de que **no hay una relación lineal exacta** de Y con X_1 y X_2, necesitamos calcular la relación residual que se expresa con **ui**; misma **que no conocemos porque pertenece al universo, razón por la cual es estimada mediante** $\boldsymbol{e_i}$, cuyo cuadrado e_i^2, minimiza la suma de cuadrados de todos los residuos: $\sum e_i^2$.

Para obtener $\boldsymbol{e_i}$ (donde (i) = 1, 2, 3... ...8, 9,10), antes necesitamos determinar $\hat{y}_i$ para cada uno de los 10 años. Así:

$$\hat{y}_1 = 5.8637 + 0.4379(1) - 0.3152(8) = 3.7800$$
$$\hat{y}_2 = 5.8637 + 0.4379(2) - 0.3152(15) = 2.0115$$
$$\hat{y}_3 = 5.8637 + 0.4379(2.5) - 0.3152(10) = 3.8065$$
$$\hat{y}_4 = 5.8637 + 0.4379(3) - 0.3152(9) = 4.3406$$
$$\hat{y}_5 = 5.8637 + 0.4379(4) - 0.3152(7) = 5.4089$$
$$\hat{y}_6 = 5.8637 + 0.4379(5) - 0.3152(6) = 6.1620$$
$$\hat{y}_7 = 5.8637 + 0.4379(7) - 0.3152(8) = 6.4074$$
$$\hat{y}_8 = 5.8637 + 0.4379(8) - 0.3152(4) = 8.1061$$
$$\hat{y}_9 = 5.8637 + 0.4379(9) - 0.3152(3) = 8.8592$$
$$\hat{y}_{10} = 5.8637 + 0.4379(15) - 0.3152(1) = 12.1170$$

Nota: estos resultados están "redondeados"; la computadora ofrece más decimales; por ello las cifras no coinciden exactamente con las que se obtienen al correr los datos en computadora.

Estos resultados aparecen en la continuación de la tabla 1 en la columna integrada por $\hat{Y}_i$.

Continuación de la Tabla 1:
Relación del consumo (Y) con el ingreso (X_1) y la inflación (X_2) (del año 1 al 10)

Años	$\hat{y}_i$	**e =Y -** $\hat{y}_i$	***e***^{2}	y^2
1	3.7800	-0.7800	0.6084	9.61
2	2.0115	-0.0115	0.0001	16.81
3	3.8065	0.1936	0.0375	4.41
4	4.3406	0.6594	0.4348	1.21
5	5.4089	-0.4089	0.1672	1.21
6	6.1620	0.8380	0.7022	0.81
7	6.4074	-0.4074	0.1660	0.01
8	8.1061	-0.1061	0.0113	3.61
9	8.8592	0.1408	0.0198	8.41
10	12.1170	-0.1170	0.0137	34.81
$\sum$	61.0000	0.0000*	2.1610	80.9

*** Existen pequeñas diferencias debido a la magnitud de los decimales. La suma técnicamente da cero.**

Observe que la suma de $\sum \hat{y}_i = 61$ *es la misma que* $\sum$ *Y = 61 y por ello la suma de* $\sum e = Y - \hat{y} = 0$*; sin embargo, para probar que su suma es un mínimo elevados al cuadrado:* $\sum$***e***^{2} *= 2.1610.* Con base en los cálculos anteriores es posible obtener $e_i = Y - \hat{y}$ y la $\sum e^2 = (Y - \hat{y})^2 = 2.1610$ en la forma que aparecen en las columnas correspondientes en la continuación de la tabla 1.

Como se observa la $\sum (Y - \hat{y}) = 0$, mientras que $\sum (Y - \hat{y})^2 = 2.1610$ esto proviene de la solución de:

$$e_1 = 3 - 3.7800 = -0.7800; e_1^2 = 0.6084$$
$$e_2 = 2 - 2.0115 = -0.0115; e_1^2 = 0.0001$$
$$e_3 = 4 - 3.8065 = 0.1936; e_1^2 = 0.0375$$
$$e_4 = 5 - 4.3406 = 0.6594; e_1^2 = 0.4348$$
$$e_5 = 5 - 5.4089 = -0.4089; e_1^2 = 0.1672$$
$$e_6 = 7 - 6.1620 = +0.8380; e_1^2 = 0.7022$$
$$e_7 = 6 - 6.4074 = -0.4074; e_1^2 = 0.1660$$
$$e_8 = 8 - 8.1016 = -0.10160; e_1^2 = 0.0113$$
$$e_9 = 9 - 8.8592 = +0.1408; e_1^2 = 0.0198$$
$$e_{10} = 12 - 12.1170 = -0.1170; e_1^2 = 0.0137$$

$$\sum e^2 = 2.1610$$

V.2.3.- Verificación de la Teoría Económica con la prueba de Significación Estadística de los parámetros Poblacionales.

Con estos valores podemos probar la significación estadística de los parámetros poblacionales β, π, calculando primero las varianzas de los estimadores. Las fórmulas de las varianzas son:

$$Var\ \hat{b} = \pm_u^2 \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

$$Var\ \hat{c} = \pm_u^2 \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2}$$

Conviene decir que con base en la forma como está planteada esta teoría económica, aquí no se calcula la varianza de $\hat{a}$ (ordenada al origen) por que no ayuda a probar la teoría económica, ya que lo que se desea es probar **las variaciones** de $\hat{y}_i$ en función de las variaciones de X_1, X_2.

Por otra parte, como $\pm_u^2$ es desconocida por que es del universo, se estima con S^2, como una estimación insesgada, es decir, E (S^2) = $\pm_u^2$.
Su fórmula es:

$$S^2 = \hat{\pm}_u^2 = \frac{\sum e_i^2}{n-k}$$

Dónde:

n = número de términos (10 en este ejemplo)
k = número de parámetros estimados (3 en este ejemplo)

Por lo que el número de los **grados de libertad** es de: 10 - 3 = **7**

Con estas referencias, las estimaciones insesgadas de las varianzas de b, c tendrán las siguientes fórmulas:

$$S_{\hat{b}}^2 = Var(\hat{b}) = \frac{\sum e_i^2}{n-k} \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} ; S_{\hat{b}} = \sqrt{S_{\hat{b}}^2}$$

$$S_{\hat{c}}^2 = Var(\hat{c}) = \frac{\sum e_i^2}{n-k} \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} ; S_{\hat{c}} = \sqrt{S_{\hat{c}}^2}$$

$S_{\hat{b}}$, $S_{\hat{c}}$ son los errores estándar de $\hat{b}, \hat{c}$ con los que se pueden probar las hipótesis sobre β, π, relativos a que hay o no hay relación de Y con respecto a X_1, X_2, con la estadística *t* de Student porque n < 30 términos:

Para β:
H_o: β = 0
H_a: β≠ 0

$$t = \frac{\hat{b} -}{S_{\hat{b}}}$$

Para π:
H_o: π= 0
H_a: π≠ 0

$$t_{®} = \frac{\hat{c} - ®}{S_{\hat{c}}}$$

Estas *t´s* “empíricas” u observadas se contrastan con las *t´s* “teóricas” o de tablas. Estas últimas se denotan con las literales t_6 , donde ∝ = nivel de significación que junto con los grados de libertad (n-k) determinan los puntos “críticos” donde se toma la decisión de aceptar o rechazar H_o.

Así, para probar H_o primero calculamos $S_{\hat{b}}^2$, $S_{\hat{c}}^2$; sustituyendo los valores de las tablas en las fórmulas originales:

$$S_{\hat{b}}^2 = Var(\hat{b}) = \frac{\sum e_i^2}{n-k} \frac{\sum x_2^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} = \frac{2.1610}{10-3} \frac{140.9}{(161.0250)(140.9) - (-123.15)^2} =$$

$$= 0.3087 \frac{140.9}{22{,}688.4225 - 15{,}165.9225} = \frac{140.9}{7{,}522.5} = 0.3087(0.0187) = 0.0058$$

$$; S_{\hat{b}}^2 = 0.0058; luego S_{\hat{b}} = \sqrt{0.0058} = 0.0761$$

También:

$$S_{\hat{c}}^2 = Var(\hat{c}) = \frac{\sum e_i^2}{n-k} \frac{\sum x_1^2}{\sum x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2} \frac{2.1610}{10-3} \frac{161.0250}{(161.0250)(140.9)-(-123.15)^2} =$$

$$= 0.3087 \frac{161.03}{22{,}688.4225 - 15{,}165.9225} = \frac{161.03}{7{,}522.5} = 0.3087(0.0214) = 0.0066$$

$$; S_{\hat{c}}^2 = 0.0066; luego S_{\hat{c}} = \sqrt{0.0066} = 0.0812$$

Por lo tanto las *t´s* empíricas serán:

$$t \;= \frac{\hat{b}-}{S_{\hat{b}}} = \frac{0.4379-0}{0.0761} = 5.7542 \qquad t_{®} = \frac{\hat{c}-®}{S_{\hat{c}}} = \frac{-0.3087-0}{0.0812} = -3.8017$$

Ahora se determina con ∝ = 5% y 7 grados de libertad (n-k), el valor en tablas de $t_{\sim} = \pm 2.365$. Vemos que β, y π estadísticamente son significativamente diferentes de cero.

Interpretación económica: Sí hay relación del consumo (Y) con el ingreso (X_1) y la inflación (X_2); luego entonces se sigue confirmando nuestra teoría.

V.2.4.- Determinación del Grado (o porcentaje) de la Relación que Existe entre Y y las Variables Explicativas X_1, X_2.

Para llevar a cabo la determinación del grado de la relación que existe entre la variable dependiente o explicada y las variables independientes o explicativas primero se determina el **Coeficiente de Determinación Múltiple.**

Como se vio en el MLS, recuérdese que la fórmula y el cálculo del coeficiente puede plantearse de la siguiente manera:

$$R^2 = 1 - \frac{\sum e_i^2}{\sum y_i^2} = 1 - \frac{2.1610}{80.9} = 1 - 0.0267$$

$$R^2 = 0.9733 \;\; o \;\; 97.33\%$$

Interpretación: El ingreso y la inflación determinan el 97.33% de los cambios que experimenta el consumo.

Importancia de $\overline{R}^2$: Coeficiente de Determinación ajustado.

Si tomamos en cuenta la reducción en los grados de libertad conforme aumenta el número de variables independientes (si antes suponíamos que el consumo sólo era función del ingreso, regresión simple, y si ahora hemos agregado la inflación como variable explicativa adicional), entonces se debe calcular el $\overline{R}^2$, cuya fórmula es:

$$\bar{R}^2 = 1-(1-R^2)\frac{n-1}{n-k} = 1-(1-0.9733)\frac{10-1}{10-3}$$
$$= 1-(1-0.9733)(1.2857)$$
$$= 1-(0.0267)(1.2857) = 1-0.0343$$
$$\bar{R}^2 = 0.9656$$

Interpretación: El 96.56% de los cambios que experimenta el consumo, se explican por cambios en las variables ingreso e inflación (variación explicada y, el restante 3.44%, que corresponde a la variación no explicada,

V.2.5.- Prueba de la Significación Global de la Regresión Múltiple.

En este caso la hipótesis nula se prueba con *F*, estadística que se refiere al análisis de varianza que es el cociente de dividir la varianza explicada entre la varianza no explicada. Su fórmula es:

$$F_{k-1,n-k} = \frac{\sum \hat{y}_i^2 \big/ (k-1)}{\sum e_i^2 \big/ (n-k)} = \frac{R^2 \big/ k-1}{(1-R^2) \big/ (n-k)}$$

Dónde:

(k-1) son los grados de libertad de la varianza explicada.
(n-k) los grados de libertad de la varianza no explicada.
(ya es conocido el significado de n y k)

Como en el caso anterior - cuando usamos *t* - se requiere encontrar en tablas la *“F teórica”* con un cierto valor de∝ para confrontarla con la *“F empírica*”. Así, primero calculamos la F empírica:

$$F_{2,7} = \frac{0.9733 \big/ 2}{1-0.9733 \big/ 7} = \frac{0.4866}{0.0038} = 128.0526$$

Si decimos que ∝ = 1%, buscamos $F_ó$en tablas con 2 grados de libertad para el **numerador** (varianza explicada) y 7 grados de libertad para el **denominador**, vemos que $F_ó$ = 9.55.

Como $F_{2,\ 7}$ = 128.0526 [$F_ó$= 9.55 decimos que se acepta la hipótesis alternativa (se rechaza la hipótesis nula) de que b y c, y R^2 son significativamente diferentes de cero.

Lo anterior indica que a través de una sola estadística se confirma la hipótesis que hemos venido desarrollando de que: Y = f(X_1, X_2).

V.2.6.- Coeficientes de Correlación Parcial.

Estos coeficientes son útiles por que miden la correlación parcial entre la variable dependiente (Y) y una variable independiente (sea X_1 o X_2) después de suponer que la(s) otra(s) variable(s) independiente(s) permanecen constantes en el modelo uniecuacional. Así, decimos en el caso de la siguiente notación que: r_{yx_1,x_2} es la correlación parcial entre X_1 e Y después de suponer que los cambios en X_2 no afectan a Y, Algo similar se puede decir de X_2 .Sus fórmulas son:

$$r_{yx_1,x_2} = \frac{r_{yx_1} - r_{yx_2}\, r_{x_1,x_2}}{\sqrt{1-r^2{}_{x_1,x_2}}\sqrt{1-r^2_{yx_2}}}$$

En forma análoga:

$$r_{yx_2,x_1} = \frac{r_{yx_2} - r_{yx_1}\, r_{x_1,x_2}}{\sqrt{1-r^2{}_{x_1,x_2}}\sqrt{1-r^2_{yx_2}}}$$

Así, procedemos a calcularlos con base en los datos de la tabla 1:

$$r_{yx_1} = \frac{\sum yx_1}{\sqrt{\sum x_1^2}\sqrt{\sum y^2}} = \frac{109.35}{\sqrt{161.03}\sqrt{80.9}} = \frac{109.35}{(12.68)(8.99)} = \frac{109.35}{113.99} = 0.959$$

$$r_{yx_2} = \frac{\sum yx_2}{\sqrt{\sum x_2^2}\sqrt{\sum y^2}} = \frac{-98.1}{\sqrt{140.9}\sqrt{80.9}} = \frac{-98.1}{(11.87)(8.99)} = \frac{-98.1}{106.71} = -0.919$$

$$r_{x_1x_2} = \frac{\sum x_1x_2}{\sqrt{\sum x_2^2}\sqrt{\sum x_1^2}} = \frac{-123.15}{\sqrt{140.9}\sqrt{161.03}} = \frac{-123.15}{(11.87)(12.68)} = \frac{-123.15}{150.51} = -0.818$$

Sustituyendo:

$$r_{yx_1,x_2} = \frac{r_{yx_1} - r_{yx_2}\, r_{x_1,x_2}}{\sqrt{1-r^2{}_{x_1,x_2}}\sqrt{1-r^2_{yx_2}}} = \frac{(0.959)-(-0.919)(-0.818)}{\sqrt{0.33}\sqrt{0.15}} =$$

$$= \frac{(0.959)-(0.751)}{(0.574)(0.387)} = \frac{0.208}{0.222} =$$

$$r_{yx_1,x_2} = 0.936 \; o \; 93.6\%$$

$$También:$$

$$r_{yx_2,x_1} = \frac{r_{yx_2} - r_{yx_1}\, r_{x_1,x_2}}{\sqrt{1-r^2{}_{x_1,x_2}}\sqrt{1-r^2_{yx_2}}} = \frac{(-0.919)-(0.959)(-0.818)}{\sqrt{0.33}\sqrt{0.08}} =$$

$$= \frac{-0.919+0.784}{(0.574)(0.282)} = \frac{-0.135}{0.161} = -0.838$$

$$luego\; r_{yx_2,x_1} = -0.838 \quad o\; 83.8\%$$

En este contexto, decimos que como $r_{y\,x_1,x_2} = 0.936$ " y" $r_{y\,x_2,x_1} = -0.838$

Se indica que el ingreso (X_1) es más importante explicando las variaciones del consumo (Y) que la inflación (X_2); obviamente en sentido inverso, como lo indica la teoría económica.

V.2.7.- Beneficios adicionales que da el MLG: Cálculo de la Elasticidad.

Con base en la información anterior se puede obtener:

V.2.7.1.- La Elasticidad Ingreso del Consumo.

Mide el cambio porcentual en los niveles del consumo como consecuencia de un cambio porcentual en el ingreso disponible (después de impuestos). Es importante señalar que la elasticidad no es constante, es decir, cambia en cada uno de los puntos de la función de regresión. Su cálculo puede hacerse con la fórmula:

$$E_{i\hat{b}} = \hat{b}\frac{\overline{X}_1}{\overline{Y}} = 0.44\frac{5.65}{6.1} = 0.44(0.92) = 0.40$$

V.2.7.2.- Con base en lo anterior podemos obtener la Elasticidad Inflación del Consumo así.

$$E_{i\hat{c}} = \hat{c}\frac{\overline{X}_2}{\overline{Y}} = -0.31\frac{7.1}{6.1} = -0.31(1.16) = -0.36$$

V.2.7.3.- *El uso de variables ficticias como variables explicativas en la regresión y correlación múltiple.*

Comenta el Profesor Mason et al (2000, 489) que en ocasiones es conveniente usar variables cualitativas, denominadas ficticias, dicotómicas, binarias, categóricas o dummy (en inglés), para explicar mejor la variable dependiente, también llamada regresada, endógena o explicada. Ahora bien, hasta el momento hemos usado variables cuantitativas como variables exógenas; sin embargo, una vez que se toma la decisión de también incluir variables cualitativas (*ficticia*), por ejemplo el sexo, religión, ocupación, grado de estudios, etc. *éstas* sólo tienen dos resultados, que se codifican como 1 y 0, es decir al describir una *cualidad*, se codifican con 1 si la tienen y 0 cuando no la tienen.

Derivado de lo anterior, en opinión de Gujarati (2003) las variables dicotómicas son un recurso para clasificar datos en categorías mutuamente excluyentes, como digamos el sexo se clasifica en dos categorías: femenino y masculino.

Ahora bien, cuando un modelo de regresión sólo incluye variables explicativas que son cualitativas se denominan *modelos de análisis de varianza*, **ANOVA.** Así, por ejemplo, sirven para probar si estadísticamente dos o más promedios difieren significativamente y por esa razón son una extensión de la prueba t que prueba dicha diferencia estadística sólo para dos promedios.

Cuando el modelo contiene variables explicativas que son cualitativas y cuantitativas, se llama *modelo de análisis de covarianza*, **ANCOVA,** que como puede intuirse, es una generalización de un modelo **ANOVA,** y en el que sus variables cuantitativas se denominan **covariantes.**

Lo anterior visto en el contexto de la regresión lineal indica que los modelos ANOVA se usan para evaluar la significación estadística de la relación entre una variable dependiente, que es cuantitativa, con una o varias independientes, que son cualitativas, en tanto que un modelo ANCOVA se usa para evaluar la significación estadística de la relación entre una variable dependiente, que es cuantitativa, con varias independientes, en que al menos una de ellas es cuantitativa o covariante.

Ejemplo: la Empresa "Arriba Juárez" desea calcular los costos de calefacción (Y) durante el invierno pasado y verificar si tienen dichos costos relación con la temperatura (X_1), el aislamiento térmico (X_2), y la existencia de un "garage" en las casas (X_3). Así, la variable independiente "garage" es cualitativa y se define con 0 cuando las casas no tengan garage y con 1 cuando lo tengan. Para ello toma una muestra de 20 casas y encuentra que:

Casa	Y= costo calefacción.	X_1=Temperatura grados, Farenhait	X_2=aislamiento pulgadas de protección:	X_3= garage
1	$250	35	3	0
2	360	29	4	1
3	165	36	7	0
4	43	60	6	0
5	92	65	5	0
6	200	30	5	0
7	355	10	6	1
8	290	7	10	1
9	230	21	9	0
10	120	55	2	0
11	73	54	12	0
12	205	48	5	1
13	400	20	5	1
14	320	39	4	1

15	72	60	8	0
16	272	20	5	1
17	94	58	7	0
18	190	40	8	1
19	235	27	9	0
20	139	30	7	0

Resolvió el modelo con el método MCO usando el software MINITAB para hallar el valor de los coeficientes del mismo y encontró la siguiente ecuación de regresión múltiple:

$Y= 393.67-3.9628X_1-11.334X_2+77.43X_3$

Donde 393.67=**a** ;- 3.9628=**b** ; -11.334=**c** ; 77.43=**d**

Al analizar los signos de los coeficientes o parámetros se observa que la teoría económica se cumple porque efectivamente hay una relación inversa entre Y e X_1, como también existe con X_2, pero en cambio, es positiva con X_3. Lo anterior es un buen indicio que indica *que podemos hacer el análisis deseado*.

Con estas referencias, entonces digamos por ejemplo, que se tienen dos casas iguales, una junto a la otra, en Ciudad Juárez, una tiene garage y la otra no; ambas tienen 3 pulgadas de aislamiento térmico y la temperatura media en enero fue de 20 grados farengheit. Para la casa sin garage, X_3 se sustituye por 0 en la ecuación de regresión. Así, el costo estimado de calefacción es de $280.90/mes, ya que

$Y= 393.67-3.9628(20)-11.334(3)+77.43(0)= \$ 280.90/mes$

Para la casa con garaje, X_3=1, luego

$Y= 393.67-3.9628(20)-11.334(3)+77.43(1)= \$ 358.30/mes$

Su diferencia es 358.30-280.90= $77.43 que es igual a **d**, luego se estima que el costo de calentar una casa con garage es $77.43 mayor que el costo de una casa equivalente sin garage. **Pero ¿esta diferencia es significativa estadísticamente**?

La respuesta *se* obtiene probando la hipótesis nula siguiente.

Ho: **d**=0;

Ha: **d**$\neq$0

Con α = 5% se prueba la hipótesis usando la estadística t como en los ejemplos anteriores, tal que obtuvo la t calculada (también llamada empírica en la página 490) igual a 3.40, misma que comparó con la t de tablas (también llamada teórica), la cual con ~ = 5% y n-k= 20-4 = 16 grados de libertad resultó ser igual a 2.12. Como la t empírica resultó estar fuera de la zona de aceptación de la t “teórica”, es decir, está en la zona de rechazo de que d=0, entonces vemos que el coeficiente de regresión d no es cero, aceptamos la Ha, de manera que concluimos indicando que el garage si es una variable explicativa de los costos de calefacción en las casas, por lo que $ 77.71 si es una diferencia significativa en el costo de calentar una casa en invierno en Ciudad Juárez. Luego entonces si es conveniente incluir la variable independiente X_3 en el análisis de costos de calefacción.

VII.3.1 El coeficiente de correlación por rangos de Spearman.

Dominick Salvatore (1993:105) presenta el siguiente ejemplo ilustrativo

sobre cómo medir la asociación que existe entre dos variables cualitativas.

Dados los siguientes datos

a) Hallar el rango o coeficiente de correlación de Spearman entre la nota de mitad de curso y el rango del CI (coeficiente de inteligencia) de una muestra aleatoria de 10 estudiantes de una gran clase, tal como la de la tabla siguiente.
b) ¿Cuándo se usa la correlación por rango?

Respuestas:
Con los siguientes datos:

Nota de mitad de curso y rangos de CI										
Estudiante	1	2	3	4	5	6	7	8	9	10
Nota de mitad de curso	77	78	65	84	84	88	67	92	68	96
Grado de CI	7	6	8	5	4	3	9	1	10	2

a) $r' = 1 - \frac{6\sum D^2}{n(n^2-1)}$

dónde
D= diferencia entre rangos de pares correspondientes de las dos variables (en orden ascendente o descendente, con el rango medio asignado a observaciones del mismo valor).

n= número de observaciones.

$$r' = 1 - \frac{6\sum D^2}{n(n^2-1)} = 1 - \frac{6(10.50)}{10(99)} = 1 - \frac{63}{990} \cong 0.94$$

Con ello se constata que existe una fuerte asociación entre las dos variables cualitativas.

Cálculos para hallar el coeficiente de correlación por rangos					
n	Nota de mitad de ciclo	Rango sobre medio ciclo	Grado de CI	D	D^2
1	96	1	2	-1	1
2	92	2	1	1	1
3	88	3	3	0	0
4	84	4.5	4	0.5	0.25
5	84	4.5	5	-0.5	0.25
6	78	6	6	0	0
7	77	7	7	0	0
8	68	8	10	-2	4
9	67	9	9	0	0
10	65	10	8	2	4
				$\sum D^2 =$	10.5

b) La correlación por rangos de Spearman se usa con datos cualitativos tales como profesión, educación, sexo, etc. cuando, por la ausencia de valores numéricos, no se puede encontrar el coeficiente de correlación de Pearson. La correlación por rangos también se usa cuando no se tienen disponibles valores precisos para todas o algunas de las variables (así que, una vez más, no se puede encontrar el coeficiente de correlación). Aún más, con un gran número de observaciones de valores grandes, **r’** se puede hallar como una estimación de **r** con el fin de evitar cálculos muy dispendiosos (sin embargo, el fácil acceso a las computadoras ha eliminado esta razón para usar **r’**).

V.2.8. Aplicaciones con Eviews 5.

Ejemplos utilizando EViews 5.

La ilustración de la regresión múltiple se realizará mediante el uso de EViews 5 utilizando datos actuales de la economía mexicana. **Para ello se retomarán los ejemplos del tema de regresión simple.**

<u>Formulación de la teoría económica.</u>

Ejemplo 1: Teoría keynesiana del consumo.

De acuerdo con la teoría keynesiana del consumo, éste depende del ingreso. Por lo que:

$$\widehat{Consumo} = \hat{a} + \hat{b}PIB + \hat{e}$$

Donde:

$\hat{a}$ = Consumo autónomo, mismo que debe ser mayor que cero;

$\hat{b}$ = Propensión Marginal a Consumir (PMgC), cuyo valor oscila entre cero y uno;

$\hat{e}$ = Término de error o variable estocástica;

PIB = Variable independiente; y

Consumo = Variable dependiente.

En dicho análisis todo o parte del ingreso adicional que obtienen las familias o la sociedad en su conjunto lo destinan a consumir o a ahorrar. Así, la diferencia 1- $\hat{b}$ es igual a la Propensión Marginal a Ahorrar (PMgS). Si $\hat{b}$ = 0, todo el ingreso adicional se destina al ahorro; si $\hat{b}$ = 1, todo el ingreso adicional se destina al consumo; si 0 < $\hat{b}$ < 1, entonces parte del ingreso adicional se destina al consumo y parte al ahorro.

Sin embargo, existen otros determinantes que explican el comportamiento del consumo, entre ellos la tasa de interés, ya que la sociedad o las familias recurren al crédito para consumir o incrementar su consumo, ya sea para adquirir bienes duraderos o no duraderos. La tasa de interés es el precio que se paga por el dinero prestado. Entre mayor sea ésta menor será la motivación para pedir prestado y, por tanto, para consumir, y viceversa, entre menor sea la tasa de interés, mayor será la motivación para pedir prestado y, por tanto, para consumir.

Otro factor que influye en el comportamiento del consumo es la inflación. Ésta mide el incremento sostenido y generalizado de los precios en los bienes y servicios. Así, si la inflación se incrementa el consumo disminuye ya que al incrementarse el precio de los bienes y servicios disminuye el consumo real, medido en precios, es decir, medido en términos de los bienes y servicios que se pueden comprar, y viceversa, si la inflación disminuye el consumo se incrementará.

En este contexto, el modelo econométrico planteado es el siguiente:

$$\widehat{Consumo} = \hat{a} + \hat{b}_1 * PIB - \hat{b}_2 * CETES - \hat{b}_3 * Inflación + \hat{e}$$

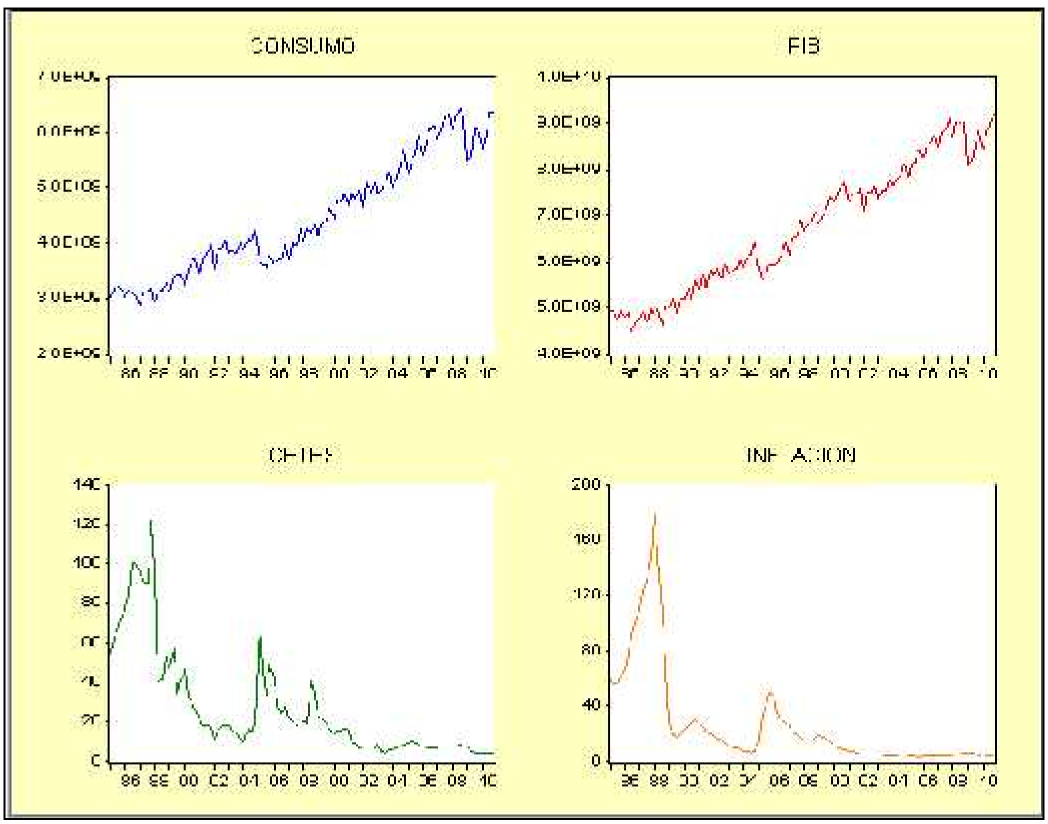
Para llevar a cabo la regresión se utilizarán datos de la economía mexicana a partir del primer trimestre de 1985 hasta el cuarto trimestre de 2010. La variable consumo corresponde al consumo privado medido en miles de pesos a precios de 2003, el ingreso corresponde al Producto Interno Bruto (PIB) expresado en miles de pesos a precios constantes de 2003, la tasa de interés corresponde a los

CETES a 28 días medida en porcentaje y la inflación corresponde a la tasa de inflación expresada en porcentaje.

Por otra parte, como se recordará, una vez que se ha generado el archivo de trabajo (Workfile) en EViews, para ingresar los datos, en la barra de comandos escribimos: “Data consumo pib cetes inflacion” seguido de la tecla enter, ello desplegará una ventana tipo Excel en la cual se copiarán los datos.

Pasemos a realizar el análisis gráfico de las series. Para ello en EViews abrimos el objeto grupo que contiene todas las series, en este caso cuatro, o seleccionamos cada una de las series y las abrimos como objeto grupo. En la ventana que se despliega vamos a View/ Multiple Graphs/ Line, con lo que se desplegará una ventana como la que se muestra en la Gráfica 1.1.

Gráfica 1.1 Análisis gráfico de las series.



Como puede apreciarse, el consumo presenta un comportamiento creciente constante, sin embargo, ha atravesado por dos claros periodos de contracción: 1) La crisis económica de 1994-1995 y, 2) La crisis económica de 2008-2009. Por lo que respecta al ingreso (PIB), éste también presenta un comportamiento creciente a lo largo del periodo de estudio y ha atravesado por las mismas dos grandes contracciones que el consumo. En cuanto a la tasa de interés (CETES a 28 días) se pueden distinguir claramente dos periodos: 1) El que va de 1985 a 1995 y, 2) El que va de 1995 a la actualidad. Para los dos casos se observa que la política o estrategia seguida para salir de la crisis de 1982 y 1995 fue primeramente incrementar la tasa de interés para atraer capitales al país y así estabilizar el tipo de cambio, hacer frente a los pasivos con el exterior y adquirir productos que no se elaboran internamente, para posteriormente disminuir la tasa de interés con la finalidad de canalizar dichos recursos a la esfera productiva, es decir, hacia la inversión. Llama la atención que tras la crisis económica de 2008-2009 no se haya incrementado significativamente la tasa de interés, ello en parte puede ser atribuible a la estabilidad macroeconómica del país y a las grandes reservas internacionales con las que cuenta el país, aunque sería preciso realizar un análisis más profundo sobre el particular. Por último, la inflación presenta un comportamiento decreciente, observándose cuatro periodos: 1) De 1985 a 1987, 2) De 1988 a 1994, 3) De 1995 a 2002 y, 4) De 2002 a 2010. En el primero se observa un comportamiento creciente de la inflación ante las crisis recurrentes en los años ochenta (1982 y 1986-1987). Para el segundo periodo la inflación disminuye como consecuencia de la reestructuración de la deuda externa y como resultado del crecimiento de la economía a partir de los pactos entre el gobierno, los empresarios y los trabajadores (contención salarial). En el tercer periodo la inflación vuelve a dispararse ante la crisis económica de 1994-1995 y comienza a disminuir como resultado del mejor desempeño de la economía acompañado de una política monetaria restrictiva y de contención salarial. El cuarto periodo puede considerarse como de estabilidad en la inflación ya que ésta ha permanecido constante y no ha sobrepasado de un dígito en los últimos años. Nuevamente, llama la atención que ésta no se haya incrementado ante la crisis de 2008-2009 y ello puede deberse, en parte, a la estabilidad macroeconómica del país, a la política monetaria restrictiva y a la contención salarial.

En cuanto al análisis estadístico de las series (ver Cuadro 1.1) observamos que ninguna de ellas se distribuye como una normal ya que en ninguno de los casos la media es igual a la mediana. En todos los casos las series no son simétricas y siguen una distribución asimétrica positiva o cargada a la derecha ya que el coeficiente de asimetría (Skewness) es mayor que cero. Asimismo, las series consumo y PIB son platocúrticas ya que el valor de la curtosis en los dos casos es menor a 3 y en el caso de los CETES y la inflación, al ser la curtosis mayor que 3, las series son leptocúrticas. Por lo que, ninguna de las series se distribuye normalmente.

Cuadro 1.1. Análisis estadístico de las series.

	CONSUMO	PIB	CETES	INFLACION
Mean	4.44E+09	6.73E+09	26.48731	26.08702
Median	4.15E+09	6.72E+09	10.55000	12.03500
Maximum	6.41E+09	9.24E+09	122.0400	177.4600
Minimum	2.86E+09	4.51E+09	4.120000	3.100000
Std. Dev.	1.05E+09	1.10E+09	26.81046	35.93352
Skewness	0.380240	0.116012	1.656195	2.333380
Kurtosis	1.887699	1.718533	4.971173	8.061451
Jarque-Bera	7.035048	7.349297	64.38228	205.7900
Probability	0.029680	0.025368	0.000000	0.000000
Sum	4.62E+11	7.00E+11	2754.680	2713.050
Sum Sq. Dev.	1.14E+20	2.01E+20	74033.46	132995.5
Observations	104	104	104	104

Es oportuno recalcar que la ausencia de normalidad en las series nos alerta sobre el posible incumplimiento de algunos supuestos del método de estimación al relacionarlas. Sin embargo, ello no implica que las series no puedan relacionarse mediante el análisis de regresión y correlación, pues las pruebas sobre las violaciones a los supuestos del método de mínimos cuadrados se realizan sobre la serie de los residuales obtenidos a través de la regresión y no sobre las series originales. Así, el estudiante tendrá que poner especial atención en la correcta especificación del modelo en cuanto a la forma funcional, tamaño de la muestra, omisión de variables relevantes, inclusión de variables redundantes, etc.

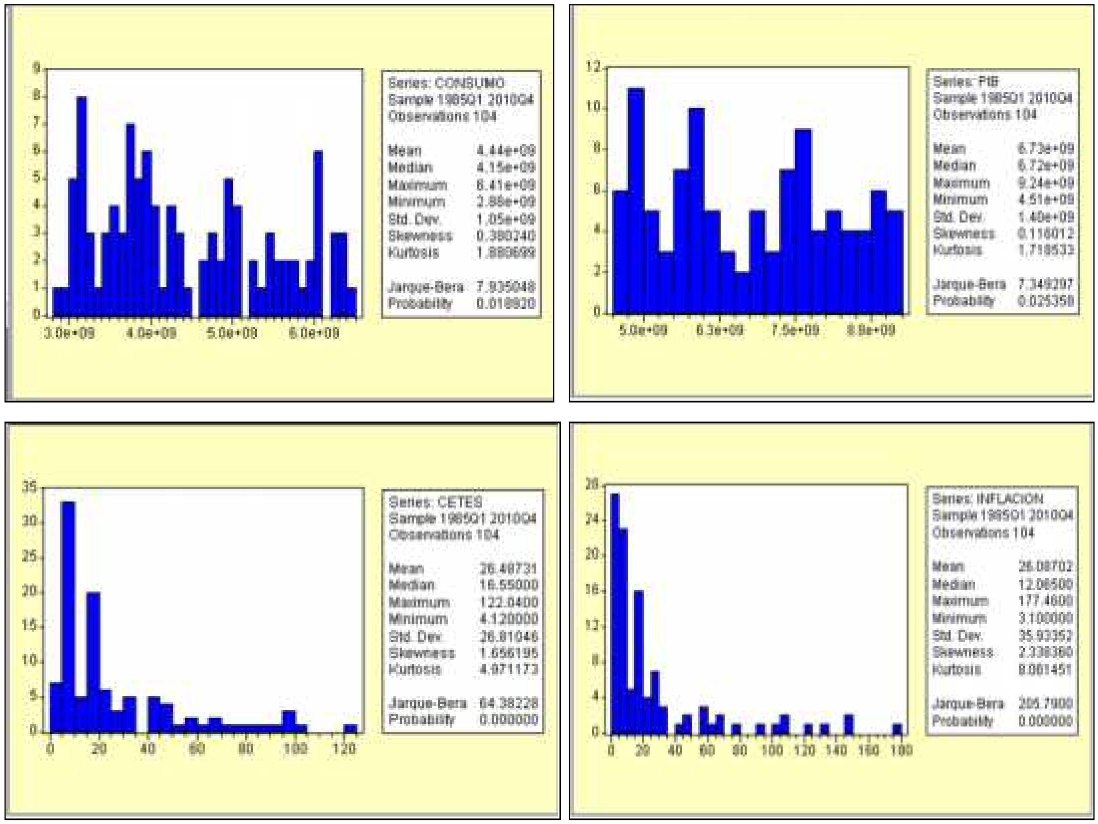
La conclusión de que las series no se distribuyen normalmente se corrobora con la estadística JB. Para ello se establece la siguiente prueba de hipótesis:

Ho: La serie se distribuye como una normal.
Ha: La serie no se distribuye como una normal.

Como se recordara, en términos ideales la estadística JB $\cong$ 6, por lo que, valores iguales o mayores a 6 en JB indican que la serie no se distribuye como una normal, en sentido contrario, valores en JB menores a 6 indican que la serie se distribuye como una normal. Por lo que en el primer caso se rechaza la Ho y en el segundo ésta no se rechaza. Asimismo, el programa proporciona una probabilidad asociada a dicha estadística que estableciendo α = 95%, indica la probabilidad mínima a la cual se rechaza la Ho. Así, si la probabilidad asociada a JB es menor a 0.05 se rechaza la Ho y si es mayor o igual a 0.05 no se rechaza dicha hipótesis.

En este sentido, en los cuatro casos se rechaza la Ho ya que la probabilidad asociada a la estadística JB es menor que 0.05, por lo que ninguna de las series se distribuye como una normal (ver Cuadro 1.1)

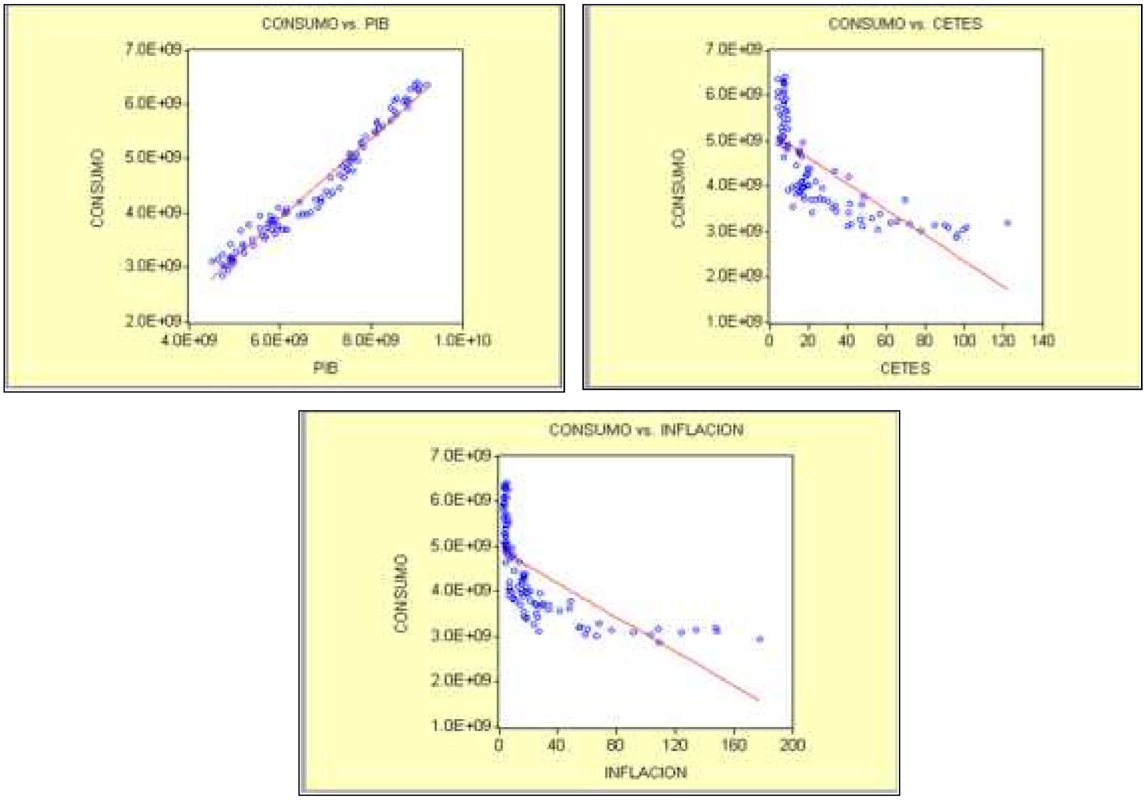
Cuadro 1.2. Histograma de las series.



El rechazo de la Ho con la estadística JB se corrobora también con el histograma de las series (ver Cuadro 1.2), pues en ninguno de los casos el histograma tiene forma de campana, es decir, en ninguno de los casos los histogramas de las variables se comportan como una normal.

Ahora realizamos los diagramas de dispersión para determinar la relación funcional que siguen cada una de las variables independientes (PIB, CETES e inflación) con respecto a la variable dependiente (consumo). En EViews para obtener el diagrama de dispersión de cada una de las variables independientes con respecto a la variable dependiente en la barra de comandos escribimos: "scat pib consumo", seguido de oprimir la tecla enter, lo mismo hacemos para las demás variables independientes. Para obtener el diagrama de dispersión con regresión lineal seleccionamos en el Workfile primero la variable independiente y después la variable dependiente, dentro de la selección damos un click derecho y vamos a Open/ As group, en la ventana de datos que se despliega vamos a View/ Graph/ Scatter/ Scatter with Regression, en la ventana que despliega no realizamos ningún cambio y oprimimos el botón de Ok., con lo que se desplegará un diagrama de dispersión como los mostrados en la gráfica 1.2

Gráfica 1.2. Diagramas de dispersión con regresión lineal.



Como puede apreciarse en la Gráfica 1.2, la relación funcional que siguen el PIB y el consumo es directa o positiva y fácilmente puede ajustarse a una línea recta, es decir, ambas siguen una relación directa lineal. En cuanto a la tasa de interés (CETES) y el consumo se observa que éstas siguen una relación inversa o negativa, sin embargo, es claro que su relación no se ajusta a un comportamiento lineal, sino más bien se ajusta a un comportamiento polinómico o inversa.

Así, a continuación procedemos a realizar la ecuación de regresión, en EViews en la barra de comandos escribimos: “ls consumo c pib cetes inflación” con lo que se obtendrá una ventana de resultados como la que se muestra en el Cuadro 1.3

Cuadro 1.3. Resultados de la ecuación de regresión.

Dependent Variable: CONSUMO
Method: Least Squares

Sample: 1985Q1 2010Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.69E+08	1.61E+08	-3.525824	0.0006
PIB	0.743448	0.020539	36.19609	0.0000
CETES	-252453.8	1579608.	-0.159821	0.8733
INFLACION	508520.0	1050962.	0.483862	0.6295

R-squared	0.966320	Mean dependent var	4.44E+09
Adjusted R-squared	0.965328	S.D. dependent var	1.05E+09
S.E. of regression	1.94E+08	Akaike info criterion	41.03999
Sum squared resid	3.75E+18	Schwarz criterion	41.14169
Log likelihood	-2130.379	F-statistic	974.3246
Durbin-Watson stat	0.614351	Prob(F-statistic)	0.000000

Por lo que la ecuación de regresión es la siguiente:

$$Con\hat{s}umo = -568786493.3 + 0.7434 * PIB - 252453.78 * CETES + 508519.97 * Inflación$$

Análisis de los signos.

De acuerdo con estos resultados, no se está cumpliendo a cabalidad con la teoría establecida, ya que $\hat{a}$ en teoría debe de ser positivo dado que representa el consumo autónomo de la sociedad, en nuestro caso éste es negativo. En cuanto al estimador de la variable inflación ($\hat{b}_3$), en teoría la relación entre ésta y el consumo es inversa, sin embargo, obtuvimos una relación directa. Por lo que respecta a los estimadores del PIB ($\hat{b}_1$) y los CETES ($\hat{b}_2$), éstos si cumplen con la teoría establecida puesto que el signo obtenido para el PIB y el consumo es positivo y el signo obtenido entre los CETES y el consumo es negativo, es decir, si el ingreso crece el consumo también lo hará y viceversa; mientras que si la tasa de interés crece el consumo disminuye y viceversa. Asimismo, en teoría $\hat{b}_1$, es decir, el estimador del PIB, oscila entre 0 y 1, por lo que éste cumple con lo establecido teóricamente.

El hecho de que algunos resultados no sean congruentes con lo establecido teóricamente puede ser atribuible a la mala especificación del modelo, principalmente en cuanto a la selección de la forma funcional, puesto que a partir de la observación de los diagramas de dispersión (ver gráfica 1.2) es claro que la relación entre los CETES y el consumo y entre la Inflación y el consumo no se ajustan a una línea recta, sino que siguen otra relación.

En cuanto a la interpretación de los estimadores obtenidos en la ecuación de regresión, independientemente de que no cumplan con la teoría establecida, tenemos que:

1. Si el PIB crece en mil pesos constantes de 2003, el consumo se incrementará en 0.7434 miles de pesos constantes de 2003, permaneciendo las demás variables independientes constantes, es decir, los CETES y la inflación; y viceversa, si el PIB disminuye en mil pesos constantes de 2003, el consumo disminuirá en 0.7434 miles de pesos constantes de 2003, permaneciendo todo lo demás constante.

2. Dado que $\hat{b}_1$= PMgC y ésta es igual a 0.7434, la PMgS=1-$\hat{b}_1$=1-0.7434=0.2566, por lo que, por cada mil pesos constantes de 2003 que la sociedad obtenga como ingreso adicional, ésta destinara 0.7434 miles de pesos constantes de 2003 a consumir y 0.2566 miles de pesos constantes de 2003 a ahorrar.

3. Si la tasa de interés (CETES) se incrementa en una unidad porcentual o en 1% el consumo disminuye en 252453.78 miles de pesos constantes de 2003, permaneciendo el PIB y la inflación constantes; y viceversa, si los CETES disminuyen en 1%, el consumo se incrementará en 252453.78 miles de pesos constantes de 2003, permaneciendo todo lo demás constante.

4. Si la tasa de inflación se incrementa en 1% el consumo se incrementará en 508519.97 miles de pesos constantes de 2003, permaneciendo el PIB y los CETES constantes; y viceversa, si la tasa de inflación disminuye en 1%, el consumo disminuirá en 508519.97 miles de pesos constantes de 2003, permaneciendo todo lo demás constante.

5. En cuanto al consumo autónomo, tenemos que si el PIB=0, los CETES=0 y la inflación=0, entonces éste ascendería a -568786493.3 miles de pesos constantes de 2003, lo cual no es lógico ya que la sociedad, independientemente del comportamiento de las variables explicativas incluidas en el modelo y de otros factores, posee un consumo mínimo para su sobrevivencia y reproducción.

Significación estadística de los parámetros poblacionales.

Por lo que respecta a la significación estadística de los parámetros, planteamos la siguiente prueba de hipótesis para el estimador del ingreso (PIB):

Ho: $_{1} = 0$, por tanto, el ingreso (PIB) no explica al consumo.

Ha: ${}_1 \neq 0$, por tanto, el ingreso (PIB) explica al consumo.

Como se recordara, el programa proporciona una probabilidad asociada al estadístico “t” (p-value), la cual representa la probabilidad mínima a la cual se rechaza la Ho. Los criterios, con un nivel de significancia al 5%, para rechazar o no la Ho con el p-value son:

1. Si la probabilidad asociada al estadístico “t” es menor o igual a 0.05 se rechaza la Ho.
2. Si la probabilidad asociada al estadístico “t” es mayor a 0.05 no se rechaza la Ho.

Por tanto, al ser la probabilidad asociada al estadístico “t” del ingreso menor a 0.05, se rechaza la Ho y por tanto el ingreso (PIB) si explica al consumo con un 95% de probabilidad de que así sea.

La prueba de hipótesis para el estimador de los CETES es:

Ho: ${}_2 = 0$, por tanto, la tasa de interés (CETES) no explica al consumo.

Ha: ${}_2 \neq 0$, por tanto, la tasa de interés (CETES) explica al consumo.

Como la probabilidad asociada al estadístico “t” de los CETES es mayor a 0.05, no se rechaza la Ho, por lo que la tasa de interés (CETES) no explica al consumo al 95% de probabilidad de que así sea. Dicha conclusión puede ser atribuible a la mala especificación del modelo, es decir, a la incorrecta selección de la forma funcional entre estas dos variables.

Para el caso del estimador de la inflación tenemos:

Ho: ${}_3 = 0$, por tanto, la inflación no explica al consumo.

Ha: ${}_3 \neq 0$, por tanto, la inflación explica al consumo.

Por lo que al ser la probabilidad asociada al estadístico “t” de la inflación mayor a 0.05, no se rechaza la Ho y por tanto, la inflación no explica al consumo al 95% de probabilidad de que así sea.

En cuanto a la prueba de hipótesis para el consumo autónomo, tenemos:

Ho: $\alpha = 0$, por tanto, su valor no tiene o no es de relevancia económica.
Ha: $\alpha > 0$, por tanto, su valor si es de relevancia económica.
Así, al ser la probabilidad asociada a la estadística “t” del estimador $\hat{\alpha}$ menor a 0.05, se rechaza la Ho, y por tanto, el valor del consumo autónomo es de relevancia económica. Sin embargo, el valor que asume es negativo, lo cual no

cumple con la teoría y la lógica económica. Ello también puede ser atribuible a la mala especificación del modelo.

Por otra parte, por tratarse de regresión múltiple procedemos a realizar la prueba de significancia global del modelo mediante la estadística “F”, para ello planteamos la siguiente prueba de hipótesis conjunta:

Ho: ${}_1 = {}_2 = {}_3 = 0$, por lo que, el PIB, los CETES y la inflación no explican al consumo.

Ha: ${}_1 \neq {}_2 \neq {}_3 \neq 0$, por lo que, el PIB, los CETES y la inflación si explican al consumo.

La estadística “F” viene dada por:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

Donde:

k-1 = Grados de libertad en el numerador.

n-k = Grados de libertad en el denominador.

k = Número de coeficientes estimados.

n = Tamaño de la muestra.

R^2 = Coeficiente de determinación.

Así, para nuestro caso, k=4, n=104 y R^2=0.966920, por lo que:

$$F = \frac{0.966920/(4-1)}{(1-0.966920)/(104-4)}$$

$$F = \frac{0.966920/3}{0.03308/100} = \frac{100(0.966920)}{3(0.03308)}$$

$$F = \frac{96.692}{0.09924} = 974.3246$$

Obsérvese que el valor obtenido de “F” es el mismo que proporciona el programa en la ventana de los resultados de la ecuación de regresión en la parte inferior derecha (ver Cuadro 1.3).

El valor obtenido o calculado de “F” se contrasta con la “F_α”, también conocida como estadística “F” teórica o de tablas. Para encontrar el valor de la “F” teórica se deben especificar el nivel de significación (α), en este caso α=5%, y los grados de libertad para el numerador y el denominador. Los grados de libertad en el numerador se determinan con k-1=4-1=3, y los grados de libertad en el denominador con n-k=104-4=100.

Así, la estadística "F" teórica es:

$$F_{3,100} = 2.6955$$

La estadística "F" calculada se contrasta con el valor de la estadística "F" teórica, para ello se establecen los siguientes criterios de decisión:

1. Si "F" calculada es menor que la "F" teórica no se rechaza la Ho.
2. Si "F" calculada es mayor o igual a "F" teórica se rechaza la Ho.

Como "F" calculada es mayor que la "F" teórica, 974.3246 > 2.6955, se rechaza la Ho, y por tanto, concluimos que el ingreso (PIB), la tasa de interés (CETES) y la inflación si explican al consumo.

Este resultado puede ser corroborado mediante la probabilidad asociada a la estadística "F" que el programa proporciona (p-value). Ésta indica la probabilidad mínima a la cual se rechazaría la Ho. Así, los criterios de decisión son:

1. Si la probabilidad asociada a la estadística "F" es menor o igual a 0.05 se rechaza la Ho.
2. Si la probabilidad asociada a la estadística "F" es mayor a 0.05 no se rechaza la Ho.

Por lo que, al ser la probabilidad asociada a la estadística "F" menor a 0.05 (ver Cuadro 1.5) se rechaza la Ho y se corrobora la conclusión de que las tres variables explicativas incluidas en el modelo explican a la variable explicada, en este caso, al consumo.

Dicha conclusión, a partir de la prueba de hipótesis conjunta del modelo, es incongruente con el resultado de las pruebas de significancia individual, ya que con éstas últimas se concluía que tanto la tasa de interés (CETES) como la inflación no tenían relación, de forma individual, con el consumo. Sin embargo, de forma conjunta, las tres variables incluidas en el modelo si explican a la variable dependiente. Esta incongruencia puede ser atribuible a la incorrecta especificación de la forma funcional entre las variables CETES e inflación con la variable consumo. Asimismo, este resultado nos alerta sobre la posibilidad del incumplimiento de alguno de los supuestos bajo los cuales está construido el método de estimación, temas que se tratarán más adelante

En cuanto a la prueba de bondad de ajuste tenemos que: R^2 = 0.966920 = 96.6920%, lo que significa que el 96.6920% de los cambios o variaciones en el consumo de la sociedad se explican por cambios o variaciones en el ingreso (PIB), la tasa de interés (CETES) y la inflación. Mientras que $\bar{R}^2$=0.965928 =

96.5928%, lo que significa que el 96.5928% de los cambios en el consumo se explican por cambios en las tres variables incluidas en el modelo y una variable adicional, por tanto, el 96.5928% de los cambios en el consumo se explican por cambios en cuatro variables independientes incluidas en el modelo. Como se puede infiere de este resultado, incluir una variable adicional en el modelo no repercute significativamente en el ajuste del mismo.

Ahora procedemos a analizar los coeficientes de correlación parciales mediante la matriz de correlaciones, para ello en EViews en la barra de comandos escribimos: “cor consumo pib cetes inflación”, seguido de oprimir la tecla enter, con ello se desplegará una ventana como la que se muestra en el Cuadro 1.4.

Cuadro 1.4. Matriz de correlaciones.

Correlation Matrix

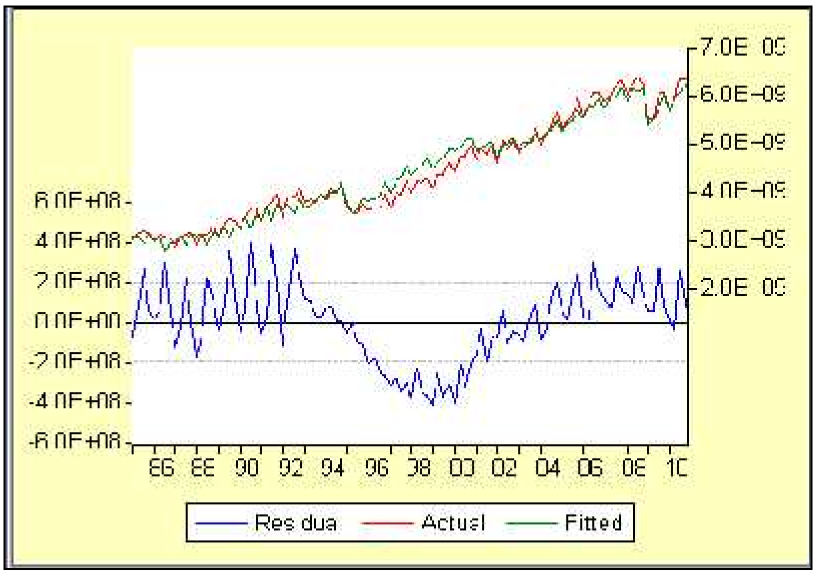
	CONSUMO	PIB	CETES	INFLACION
CONSUMO	1.000000	0.983267	-0.729654	-0.646076
PIB	0.983267	1.000000	-0.745650	-0.664552
CETES	-0.729654	-0.745650	1.000000	0.862166
INFLACION	-0.646076	-0.664552	0.862166	1.000000

Como se recordará, en la matriz de correlaciones lo que interesa analizar son los coeficientes de correlación que están por encima o por debajo de la diagonal principal de la misma, y además sólo aquellos que relacionan a la variable dependiente con las independientes y no aquellos que relacionan a éstas últimas entre ellas. Asimismo, en teoría se considerará un “r” aceptable cuando éste asume un valor mayor o igual a 0.5 en términos absolutos, trabajando con los valores se las series en sus niveles originales. Así, los coeficientes de correlación (r) que se analizarán son: el consumo con el PIB o viceversa, el consumo con los cetes o viceversa y el consumo con la inflación o viceversa.

En cuanto al coeficiente de correlación entre el consumo y el PIB observamos que este asciende a 0.983267, por lo que puede concluirse que existe una alta correlación entre estas dos series y que además su relación es directa. En el caso del consumo y los CETES se obtuvo un r=-0.729654, por lo que existe una aceptable correlación entre las dos series y su relación es inversa, como se desprende del diagrama de dispersión y tal como se obtuvo en la ecuación de regresión. Por lo que respecta al coeficiente de correlación entre el consumo y la inflación, éste fue de -0.646076, lo que significa que existe una aceptable correlación entre estas dos variables y que su relación es inversa, tal y como se observa en el diagrama de dispersión, sin embargo, el signo obtenido en la ecuación de regresión no coincide con el signo del “r” y se debe al uso de una incorrecta forma funcional entre estas dos variables.

Finalmente, a continuación se muestra gráficamente la estimación del consumo mediante la ecuación de regresión obtenida, su comparación con los datos reales de la misma variable y sus errores, es decir, su diferencia. Para ello, como se recordará, en EViews en la ventana de la regresión vamos a: View/ Actual, Fitted, Residual/ Actual, Fitted, Residual Graph, con lo cual se despliega una ventana como la que se muestra en la Gráfica 1.3.

Gráfica 1.3. Representación de los valores reales y estimados de la variable dependiente y los errores.



Ejemplo 2: Hipótesis del crecimiento económico impulsado por las exportaciones.

En consonancia con el enfoque didáctico de ampliar los tres ejemplos del MLS al MLG, ahora recordando que en el ejemplo 2 de regresión simple se planteó la hipótesis que sostiene que el crecimiento del Producto Interno Bruto (PIB) es impulsado únicamente por las Exportaciones (X). Sin embargo, no sólo las Exportaciones impulsan el crecimiento del producto, existen otras variables que afectan a éste. Entre ellas se encuentran las Importaciones (M), éstas inciden de manera negativa ya que son bienes elaborados en el extranjero aunque sean consumidos en el país, además de que representan una salida de ingresos, pues se debe pagar por ellas. En cuanto a la Inflación, como se mencionó en el ejemplo 1de regresión múltiple, mide el crecimiento de precios en bienes y servicios; en resumen, si se da un incremento en la inflación el Producto Interno Bruto disminuye, y viceversa. Por último, el Tipo de Cambio, que es el precio de una unidad de moneda extranjera reflejada en moneda nacional, afecta a las variables Importaciones y Exportaciones, dando pauta a un incremento o decremento de la competitividad, y por ende afectando directamente al Producto (PIB).

En este contexto, el modelo econométrico se plantea de la siguiente manera:

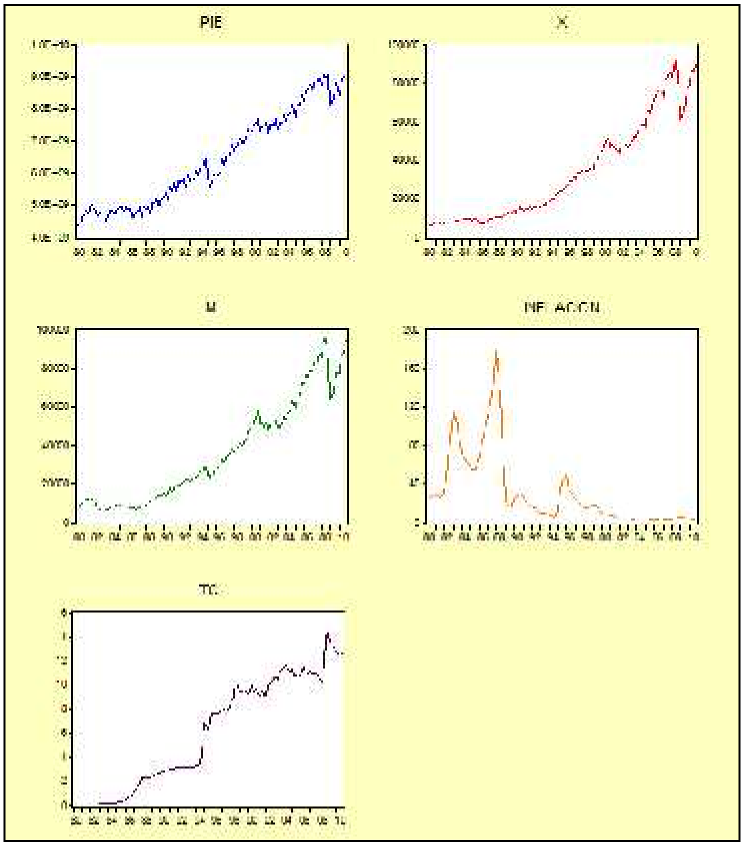
$$P\hat{I}B = \hat{a} + \hat{b}_1 * X - \hat{b}_2 * M - \hat{b}_3 * Inflación + \hat{b}_4 * TC + \hat{e}$$

Para este ejercicio de regresión múltiple, se utilizarán datos de la economía mexicana a partir del primer trimestre de 1980 hasta el cuarto trimestre de 2010. Por su parte, el Producto Interno Bruto (PIB) está expresado en miles de pesos a precios constantes de 2003, las Importaciones y Exportaciones están expresadas en millones de dólares, la Inflación está expresada en porcentaje y el Tipo de Cambio (TC) está expresado en pesos por dólar.

A continuación pasamos a analizar el comportamiento gráfico de las series (ver gráfica 2.1). En cuanto al Producto Interno Bruto, las Exportaciones, las Importaciones y el Tipo de Cambio, éstos presentan un comportamiento creciente constante, sin embargo, éste último antes de 1994 crecía lentamente, ya que durante ese año la Comisión de Cambios que se integra por la Secretaría de Hacienda y Crédito Público y el Banco de México, acordó que el tipo de cambio se determinaría por las libres fuerzas del mercado. En tanto, la inflación durante la década de los ochenta y principios de los noventa crecía con rapidez, durante el inicio del siglo XXI decrece y se mantiene constante, ya que el objetivo del Banco de México ha sido mantener la inflación baja y estable, con el fin de establecer condiciones que favorezcan el crecimiento económico sostenido y la generación de empleos.

En cuanto al análisis estadístico de las series (ver Cuadro 2.1) observamos que ninguna de ellas se distribuye como una normal ya que en ninguno de los casos la media es igual a la mediana. En todos los casos las series no son simétricas y siguen una distribución asimétrica positiva o cargada a la derecha ya que el coeficiente de asimetría es mayor que cero. Asimismo, las series del Producto Interno Bruto, Exportaciones, Importaciones y Tipo de Cambio son platocúrticas ya que el valor de la curtosis en los cuatro casos es menor a 3 y en el caso de la Inflación, al ser la curtosis mayor que 3, la serie es leptocúrtica. Con ello se puede concluir que ninguna de las series se distribuye normalmente.

Gráfica 2.1. Análisis gráfico de las series.



La conclusión de que las series no se distribuyen normalmente se corrobora con la estadística JB. Para ello se establece la siguiente prueba de hipótesis:

Ho: La serie se distribuye como una normal.
Ha: La serie no se distribuye como una normal.

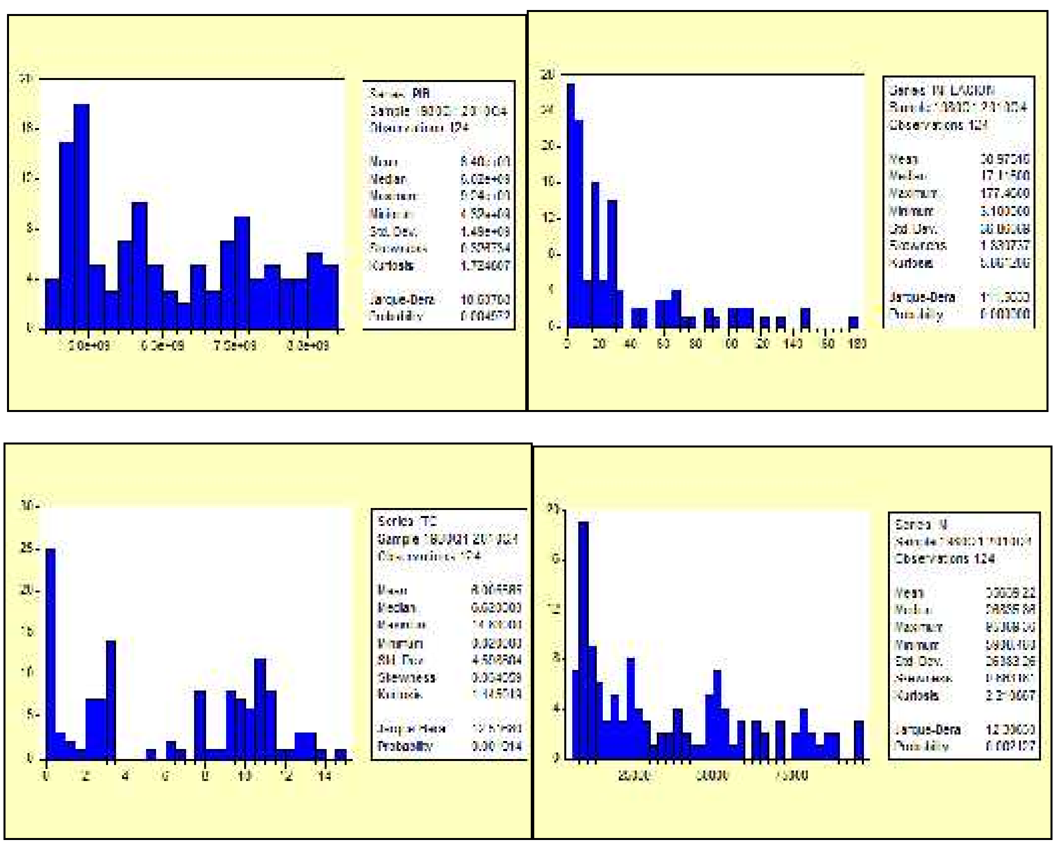
Como se recordará, para no rechazar la hipótesis nula, la probabilidad asociada al estadístico JB debe de ser mayor o igual a 0.05. Por lo que en los cinco casos se rechaza la Ho ya que la probabilidad asociada a la estadística JB es menor que 0.05, por lo que se concluye que ninguna de las series se distribuye como una normal (ver Cuadro 2.1).

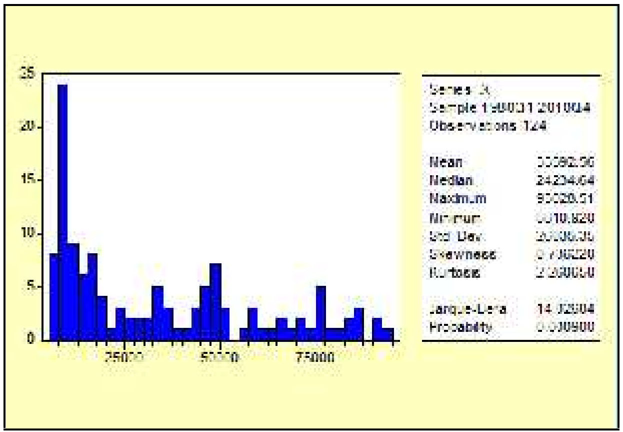
La no normalidad de las series se corrobora con sus respectivos histogramas. Para este ejemplo en ninguno de los casos el histograma tiene forma de campana, es decir, sigue confirmándose que ninguna de las series se comporta como una normal (ver Cuadro 2.2)

Cuadro 2.1. Análisis estadístico de las series.

	PIE	X	M	INFLACION	TC
Mean	6.40E+09	33392.56	35639.22	30.97516	6.005565
Median	6.02E+09	24234.64	26835.86	17.11500	6.620000
Maximum	9.24E+09	93023.51	95369.36	177.4600	14.63000
Minimum	4.32E+09	5810.920	5908.460	3.100000	0.020000
Std. Dev.	1.49E+09	26035.35	26383.26	36.86069	4.598804
Skewness	0.326734	0.736220	0.663161	1.830737	0.034059
Kurtosis	1.724807	2.260650	2.210867	5.861206	1.445019
Jarque-Bera	10.60788	14.02604	12.30630	111.5633	12.51680
Probability	0.004972	0.000900	0.002127	0.000000	0.001914
Sum	7.93E+11	4140677.	4419263.	3840.920	744.6900
Sum Sq. Dev.	2.72E+20	8.34E+10	8.56E+10	167121.4	2601.327
Observations	124	124	124	124	124

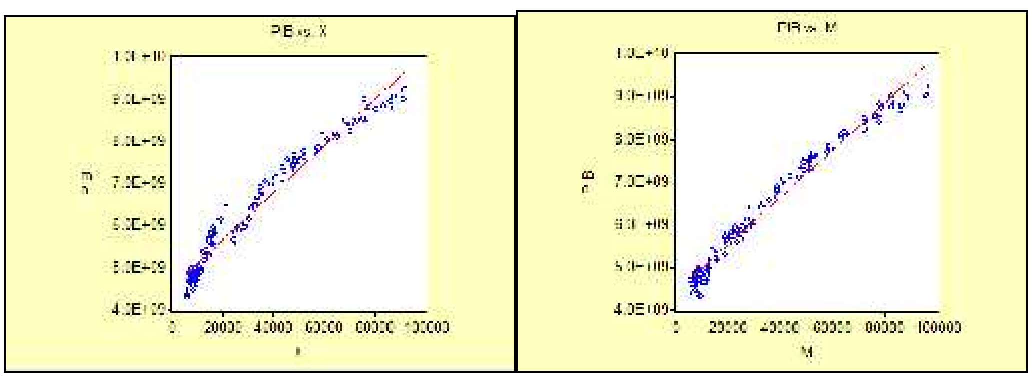
Cuadro 2.2. Histograma de las series

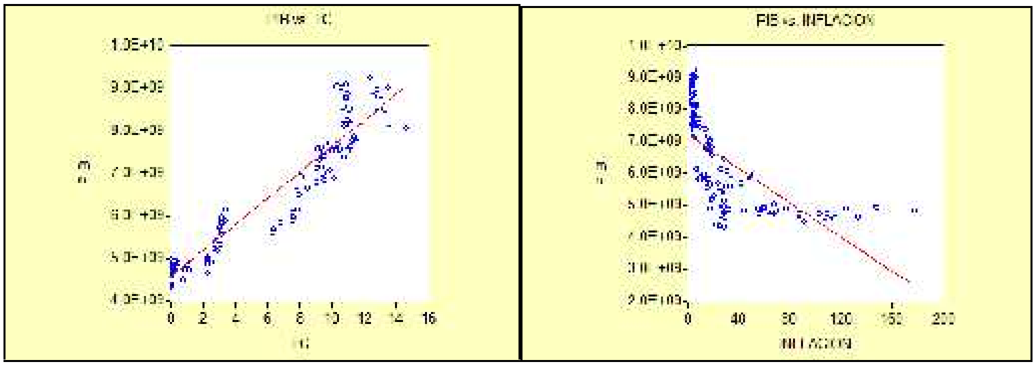




Ahora pasamos a determinar la relación que siguen cada una de las variables independientes con la dependiente. Así, en los diagramas de dispersión (ver Gráfica 2.2) puede observarse que la relación funcional que sigue el PIB con respecto a las Exportaciones se ajusta a una la línea recta y además siguen una relación directa o positiva. En cuanto a la relación que sigue el PIB con respecto a las importaciones también se ajusta a una la línea recta y su relación es directa. Por lo que respecta a la relación entre el PIB y el Tipo de Cambio ésta no se ajusta completamente a una línea recta, sin embargo, parecen seguir una relación directa. Por último, la relación entre el PIB y la Inflación no se ajusta claramente a una línea recta y la relación entre ambas variables es inversa.

Gráfica 2.2. Diagramas de dispersión lineal.





En cuanto a los resultados de la ecuación de regresión (ver Cuadro 2.3) tenemos que la ecuación de regresión es la siguiente:

$$P\hat{I}B = 4367687566 - 29026.12614 * X + 69365.26982 * M - 413171.2236 * Inflación + 89900775.1 * TC$$

De acuerdo con estos resultados, no se está cumpliendo con la teoría establecida, ya que las Exportaciones en teoría debían de mantener una relación positiva con el PIB ya que lo impulsan, al correrse el modelo el signo que se obtiene es el contrario al esperado. Asimismo, las Importaciones deberían mantener una relación negativa y en la ecuación de regresión se presenta una relación directa. El resto de las variables independientes como lo son la Inflación y el Tipo de Cambio sí mantienen las relaciones que se habían establecido en teoría con respecto al PIB. El hecho de que algunos resultados, es decir, los signos, no sean congruentes con lo establecido teóricamente puede ser atribuible a la mala especificación del modelo, ya que no todas las variables se ajustan a un comportamiento lineal (ver Gráfica 2.2 de los diagramas de dispersión e interpretación al respecto).

Cuadro 2.3. Resultados de la ecuación de regresión.

Dependent Variable: PIB
Method: Least Squares

Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.37E+09	56457040	77.36303	0.0000
X	-29026.13	8024.846	-3.617032	0.0004
M	69365.27	8171.126	8.489072	0.0000
INFLACION	-413171.2	657873.4	-0.628041	0.5312
TC	89900775	8296738.	10.83568	0.0000

R-squared	0.988286	Mean dependent var	6.40E+09
Adjusted R-squared	0.987892	S.D. dependent var	1.49E+09
S.E. of regression	1.64E+08	Akaike info criterion	40.70272
Sum squared resid	3.18E+18	Schwarz criterion	40.81644
Log likelihood	-2518.569	F-statistic	2509.997
Durbin-Watson stat	1.296021	Prob(F-statistic)	0.000000

En cuanto a la interpretación de los estimadores obtenidos en la ecuación de regresión, es la siguiente:

1. Si las Exportaciones crecen en un millón de dólares, el Producto Interno Bruto disminuye en 29026.12613 miles de pesos constantes de 2003; por el contrario, si las Exportaciones disminuyen en un millón de dólares, el Producto Interno Bruto aumentará en 29026.12613 miles de pesos constantes de 2003, permaneciendo las demás variables independientes constantes.

2 En cuanto a las Importaciones, si estas crecen en un millón de dólares, el Producto Interno Bruto aumentará en 69365.26982 miles de pesos constantes de 2003, de forma inversa, si las Importaciones disminuyen en un millón de dólares, el Producto Interno Bruto disminuirá en 69365.26982 miles de pesos constantes de 2003, permaneciendo todo lo demás constante.

3. En cuanto a la Inflación, si ésta se incrementa en 1% el Producto Interno Bruto se reducirá en 413171.2236 miles de pesos constantes de 2003; si la tasa de Inflación disminuye en 1%, el Producto Interno Bruto aumentará en 413171.2236 miles de pesos constantes de 2003, permaneciendo el resto de las variables independientes constantes.

4. En cuanto el Tipo de Cambio, por cada peso por dólar que éste aumente, el Producto Interno Bruto aumentará en 89900775 miles de pesos constantes de 2003; si el Tipo de Cambio disminuye en un peso por dólar, el Producto

Interno Bruto disminuirá en 89909775 miles de pesos constantes de 2003, permaneciendo sin cambios el resto de las variables independientes incluidas en el modelo.

5. En cuanto al Producto Interno Bruto independiente o autónomo, tenemos que si las variables independientes son iguales a cero, entonces el Producto asciende a 4367687566 miles de pesos constantes de 2003, ya que existen otros factores que determinan o inciden sobre éste.

A continuación se procederá a realizar la prueba de significación estadística de los estimadores. Para ello establecemos la siguiente prueba de hipótesis:

Ho: ${}_{1} = 0$, por tanto, las Exportaciones no explican al PIB.

Ha: ${}_{1} \neq 0$, por tanto, las Exportaciones explican al PIB.

En los ejemplos anteriores se menciona que el programa proporciona una probabilidad asociada al estadístico “t” (p-value). Los criterios, con un nivel de significancia al 5%, para rechazar o no la Ho con el p-value son:

1. Si la probabilidad asociada al estadístico “t” es menor o igual a 0.05 se rechaza la Ho.

2. Si la probabilidad asociada al estadístico “t” es mayor a 0.05 no se rechaza la Ho.

Por tanto, la probabilidad asociada al estadístico “t” de las Exportaciones es menor a 0.05, se rechaza la Ho y por tanto las Exportaciones si explican al PIB con un 95% de probabilidad.

La prueba de hipótesis para el estimador de las Importaciones es:

Ho: ${}_{2} = 0$, por tanto, las Importaciones no explican al PIB.

Ha: ${}_{2} \neq 0$, por tanto, las Importaciones explican al PIB.

En este caso la probabilidad asociada al estadístico “t” de las Importaciones es menor a 0.05, se rechaza la Ho, por lo que las Importaciones explica al PIB al 95% de probabilidad.

Para el caso del estimador de la Inflación tenemos:

Ho: ${}_{3} = 0$, por tanto, la Inflación no explica al PIB.

Ha: ${}_{3} \neq 0$, por tanto, la Inflación explica al PIB.

Aquí la probabilidad asociada al estadístico “t” de la Inflación es mayor a 0.05, se acepta la Ho y por tanto, la Inflación no explica al PIB al 95% de probabilidad.

Para el caso del estimador del Tipo de Cambio tenemos:

Ho: ${}_4 = 0$, por tanto, el Tipo de Cambio no explica al PIB.

Ha: ${}_4 \neq 0$, por tanto, el Tipo de Cambio explica al PIB.

Aquí la probabilidad asociada al estadístico “t” del Tipo de Cambio es menor a 0.05, se rechaza la Ho y por tanto, el Tipo de Cambio explica al PIB al 95% de probabilidad.

En cuanto a la prueba de hipótesis para el PIB independiente, tenemos:

Ho: α = 0, por tanto, su valor no tiene o no es de relevancia económica.

Ha: α > 0, por tanto, su valor si es de relevancia económica.

Así, al ser la probabilidad asociada a la estadística “t” del estimador $\hat{a}$ menor a 0.05, se rechaza la Ho, y por tanto, el valor del PIB independiente es de relevancia económica.

Por otra parte, por tratarse de regresión múltiple procedemos a realizar la prueba de significación global del modelo mediante la estadística “F”, para ello planteamos la siguiente prueba de hipótesis conjunta:

Ho: ${}_1 = {}_2 = {}_3 = {}_4 = 0$, las Exportaciones, Importaciones, la Inflación y el Tipo de Cambio no explican al PIB.

Ha: ${}_1 \neq {}_2 \neq {}_3 \neq {}_4 \neq 0$, las Exportaciones, Importaciones, la Inflación y el Tipo de Cambio si explican al PIB.

Como se recordará la estadística “F” viene dada por:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

En este ejemplo, k=5, n=124 y R^2=0.988286, por lo que:

$$F = \frac{0.988286/(5-1)}{(1-0.988286)/(124-5)}$$

$$F = \frac{0.988286/4}{0.011714/119} = \frac{0.2470715}{9.8437x10^{-5}}$$

$$F = 2509.94609$$

Este valor puede apreciarse en la ventana de los resultados de la ecuación de regresión en la parte inferior derecha (ver Cuadro 2.3).

Retomando el ejercicio anterior el valor calculado de “F” se contrastará con la “F_α”. Para encontrar el valor de la “F” se seguirá utilizando un nivel de significación de α=5%, y los grados de libertad para el numerador y el denominador. Los grados de libertad en el numerador son k-1=5-1=4, y los grados de libertad en el denominador con n-k=124-5=119.

Así, “F_α” es:

$$F_{4,119} = 2.4479$$

Recordando los criterios de decisión establecidos con anterioridad:

1. Si “F” calculada es menor que la “F” teórica no se rechaza la Ho.
2. Si “F” calculada es mayor o igual a “F” teórica se rechaza la Ho.

Como “F” calculada es mayor que la “F_α”, 2509.94609 >2.4479, se rechaza la Ho, y por tanto, concluimos que las Exportaciones, las Importaciones, la Inflación y el Tipo de Cambio si explican al PIB.

Anteriormente se mencionó que el programa proporciona una probabilidad asociada al estadístico “F” (p-value), éste es otra manera de corroborar los resultados del estadístico “F”. Los criterios de decisión son:

1. Si la probabilidad asociada a la estadística “F” es menor o igual a 0.05 se rechaza la Ho.
2. Si la probabilidad asociada a la estadística “F” es mayor a 0.05 no se rechaza la Ho.

Por lo que, al ser la probabilidad asociada a la estadística “F” menor a 0.05 se rechaza la Ho y se concluye que las cuatro variables explicativas incluidas en el modelo explican a la variable dependiente, en este caso, al PIB.

En cuanto a la prueba de bondad de ajuste tenemos que: R^2= 0.988286 = 98.8286%, lo que significa que el 98.8286% de los cambios o variaciones en el Producto Interno Bruto se explican por cambios o variaciones las Exportaciones, las Importaciones, la Inflación y el Tipo de Cambio. En tanto que el $\overline{R}^2$=0.987892 = 98.7892%, lo que significa que el 98.7892% de los cambios en el Producto Interno Bruto se explican por cambios en las cuatro variables incluidas en el modelo y una variable adicional, por tanto, el 98.7892% de los cambios en el Producto Interno Bruto se explican por cambios en cinco variables independientes incluidas en el modelo. Con lo que al incluir una variable adicional no repercute significativamente en el ajuste del modelo.

A continuación se analizarán los coeficientes de correlación parciales mediante la matriz de correlaciones, que se obtendrán por el programa de acuerdo con los pasos seguidos en (Cuadro 2.4), los ejercicios anteriores.

Los coeficientes de correlación (r) que se analizarán son: el PIB con las Exportaciones o viceversa, el PIB con las Importaciones o viceversa, el PIB con la Inflación o viceversa y el PIB con el Tipo de Cambio o viceversa.

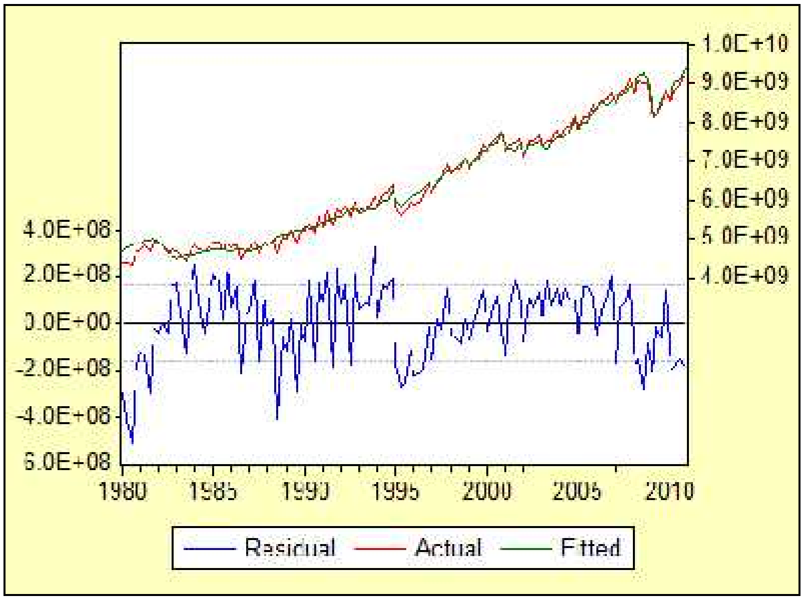
En cuanto al coeficiente de correlación entre el PIB y las Exportaciones observamos que asciende a 0.978608, por lo que existe una alta correlación entre estas dos variables y su relación es directa, tal y como se observa en el diagrama de dispersión, sin embargo, el signo obtenido en la ecuación de regresión no coincide con el signo del "r" y puede ser atribuible al uso de una incorrecta forma funcional entre el PIB y las exportaciones, así como al posible incumplimiento o violación a alguno de los supuestos bajo los cuales está construido el método de estimación . En el caso del PIB y las Importaciones el coeficiente de correlación parcial asciende a 0.986717, por lo que estas dos variables tienen una alta correlación y su relación es directa dado su signo positivo. En el caso del PIB y la Inflación el resultado que arroja es r= -0.663737 la relación entre estas dos variables el aceptable e inversa. Por último, el coeficiente de correlación entre el PIB y el Tipo de Cambio es de 0.949795, teniendo una alta correlación entre estas dos variables y mostrando una relación directa.

Cuadro 2.4. Matriz de correlaciones.

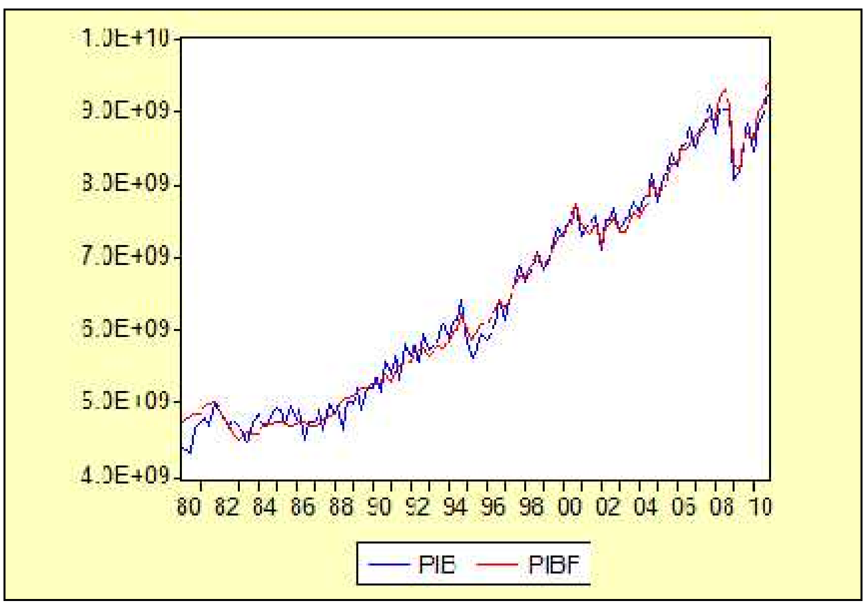
Correlation Matrix					
	PIB	X	M	INFLACION	TC
PIB	1.000000	0.978608	0.986716	-0.663737	0.949795
X	0.978608	1.000000	0.996041	-0.588187	0.915606
M	0.986716	0.996041	1.000000	0.631763	0.918392
INFLACION	-0.663737	-0.588187	-0.631763	1.000000	-0.628120
TC	0.949795	0.915606	0.918392	-0.628120	1.000000

Para finalizar este ejemplo se muestra gráficamente la estimación del PIB mediante la ecuación de regresión obtenida, su comparación con los datos reales de la misma variable y los errores de estimación (ver Gráficas 2.3 y 2.4)

Gráfica 2.3. Representación gráfica de los valores reales y estimados de la variable dependiente y los errores.



Gráfica 2.4. Gráfica del PIB y el PIB estimado.



Ejercicio 3: Teoría de la inversión.

En este ejercicio se evalúa nuevamente la teoría de la inversión en función de la tasa de interés, pero ampliando el análisis mediante la incorporación de otras variables explicativas. Dichas variables son el Producto Interno Bruto y el Consumo de Gobierno.

La explicación de la incorporación de las dos variables mencionadas, está dada por el sustento teórico neoclásico acerca de que la Inversión está relacionada negativamente con la tasa de interés, en sentido positivo con el Producto Interno Bruto y en sentido negativo con el Consumo de Gobierno.

$$I = f(Y, i, C_g)$$

Donde:
I = Inversión
i = tasa de interés
Y= Ingreso
Cg= Consumo del Gobierno

Como sabemos el Ingreso es igual al Producto Interno Bruto (PIB). En teoría macroeconómica varios son los integrantes de la inversión, el más importante son los ingresos, ya que una nueva inversión genera a la empresa ingresos adicionales si le ayuda a vender más. Eso induce a pensar que el nivel global de producción (PIB) es un importante determinante de la inversión. De hecho, la inversión depende de los ingresos que genere la situación de la actividad económica general. La inversión es muy sensible al ciclo económico.

Otro elemento fundamental que determina la inversión es el Gasto del Gobierno (Cg) sustentado en el modelo neoclásico acerca de la financiación del déficit público. En este modelo nos habla de la relación que existe entre el consumo del gobierno y el consumo privado en cuanto al stock de capital fijo. Cuando el gobierno decide endeudarse en algún programa está compitiendo con la capacidad de los empresarios para invertir, es decir, el gobierno ahora desempeña más actividades productivas o incide con un subsidio a cierto sector. Por ello le quita demanda a los productos resultantes de procesos productivos llevados a cabo por empresarios, esto da por resultado que las generaciones futuras vean reducidos sus ingresos reales y la productividad. Resultado de esto, el Estado aumenta su demanda de crédito, es decir, se eleva el precio del crédito (interés), llevando a los empresarios a reducir la inversión frenando el stock de capital para las generaciones futuras, por ello se manifiesta que el incremento del gasto público reduce la inversión en gran cuantía.

Por lo mencionado anteriormente, nuestra función esta expresada de la siguiente manera:

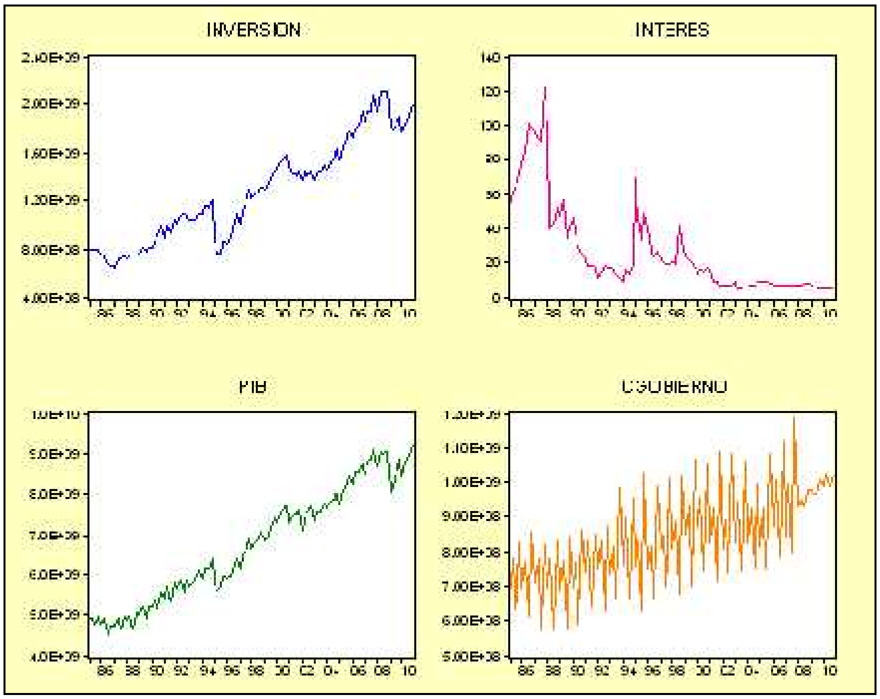
$$Inversión = \ddot{a} - \ddot{b}_1 * Interés + \ddot{b}_2 * PIB - \ddot{b}_3 * Gasto\ Público + \varepsilon$$

Para llevar a cabo la regresión se utilizarán datos de la economía mexicana a partir del primer trimestre de 1985 hasta el cuarto trimestre de 2010. Las variables Inversión y Consumo de Gobierno están medidas en miles de pesos a precios de

2003, el ingreso corresponde al Producto Interno Bruto (PIB) expresado en miles de pesos a precios constantes de 2003 y la Tasa de Interés corresponde a los CETES a 28 días medida en porcentaje.

Para comenzar el análisis de nuestra regresión es preciso observar con anticipación la tendencia de cada variable por separado, esto lo observamos en la Gráfica 3.1 que a continuación se muestra.

Gráfica 3.1. Análisis gráfico de las series.



Al analizar el Cuadro anterior tenemos que la Inversión por su parte tiene tendencia creciente, al paso del tiempo tiene 3 caídas notables. Por su parte la Tasa de Interés cuenta con una tendencia hacia la baja, hacia el año 2010 es constante y estable con las cifras más bajas en el periodo establecido para nuestro estudio, aunque se pueden observar claramente tres periodos con un aumento considerable de la tasa de interés, estos valores altos coinciden con la tendencia a la baja del interés. Se reduce el análisis de estas dos variables ya que se ha hablado de ellas en el ejercicio 3 de regresión simple.

En cuanto a nuestras nuevas variables independientes incorporadas al modelo, el PIB muestra una tendencia creciente muy parecida al comportamiento de la Inversión, sin embargo, ésta es divergente en cuanto a estabilidad, ya que el Producto Interno Bruto es más estable, es decir, no tiene bajas tan abruptas, sus bajas coinciden con los periodos de crisis por lo que ha atravesado el país. Así, antes de 1982 México había sostenido un nivel de crecimiento continuo, para

1982 comienza una crisis, ésta freno considerablemente el ritmo de crecimiento de la economía nacional ya que las tasas de interés bajas por ajuste se elevaron dado el uso de ingresos en consumo mayores que en producción, este hecho afectó la economía llegando a la situación de que México no pudo hacer frente a sus obligaciones contraídas, lo que desató la crisis de la deuda externa. Durante este periodo las tasas de crecimiento del PIB fueron muy bajas pero se logró un repunte en 1984 para que en el año de 1986 se incurriera en una crisis de hiperinflación con la cual se tuvo una notable caída que afecto fuertemente al desarrollo del país. Al llegar el año de 1987 México logra tener una tasa de crecimiento económico positiva, la cual se mantuvo relativamente estable hasta el año de 1994. Sin embargo, tras la crisis de 1994-1995 el PIB cayó casi un 8%, al recuperarse de esta crisis México vuelve a tener un ritmo de crecimiento estable llegando a tener tasas de crecimiento hasta del 5% que se vieron realmente afectadas al llegar el año 2008, año en que inicio la crisis hipotecaria estadounidense que arrastró a todas las economías. En lo que respecta al Consumo del Gobierno, en el periodo que analizamos este es muy variable pero tiene una tendencia positiva, parecida en altas y bajas igualmente con la Inversión y el PIB, lo que muestra que con las crisis el gasto se contrae para tener mayores ingresos y poder solventar la desventaja o mala situación que conlleva una crisis.

Ahora pasamos al análisis cuantitativo con el Cuadro 3.1 que muestra los estadísticos descriptivos básicos de cada variable.

Cuadro 3.1. Análisis estadístico de las series.

	INVERSION	INTERES	PIB	CGOBIERNO
Mean	1.27E+09	26.48731	6.73E+09	8.42E+08
Median	1.24E+09	16.50000	6.72E+09	8.29E+08
Maximum	2.11E+09	122.0400	9.24E+09	1.19E+09
Minimum	6.38E+08	4.120000	4.51E+09	5.72E+08
Std. Dev.	4.22E+08	26.31046	1.40E+09	1.41E+08
Skewness	0.311151	1.656195	0.116012	0.031726
Kurtosis	1.931008	4.971173	1.716533	2.234162
Jarque-Bera	6.630013	64.36228	7.349297	2.275823
Probability	0.036334	0.000000	0.025350	0.320000
Sum	1.32E+11	2754.680	7.00E+11	8.76E+10
Sum Sq. Dev.	1.83E+19	71036.46	2.01E+20	2.06E+18
Observations	104	104	104	104

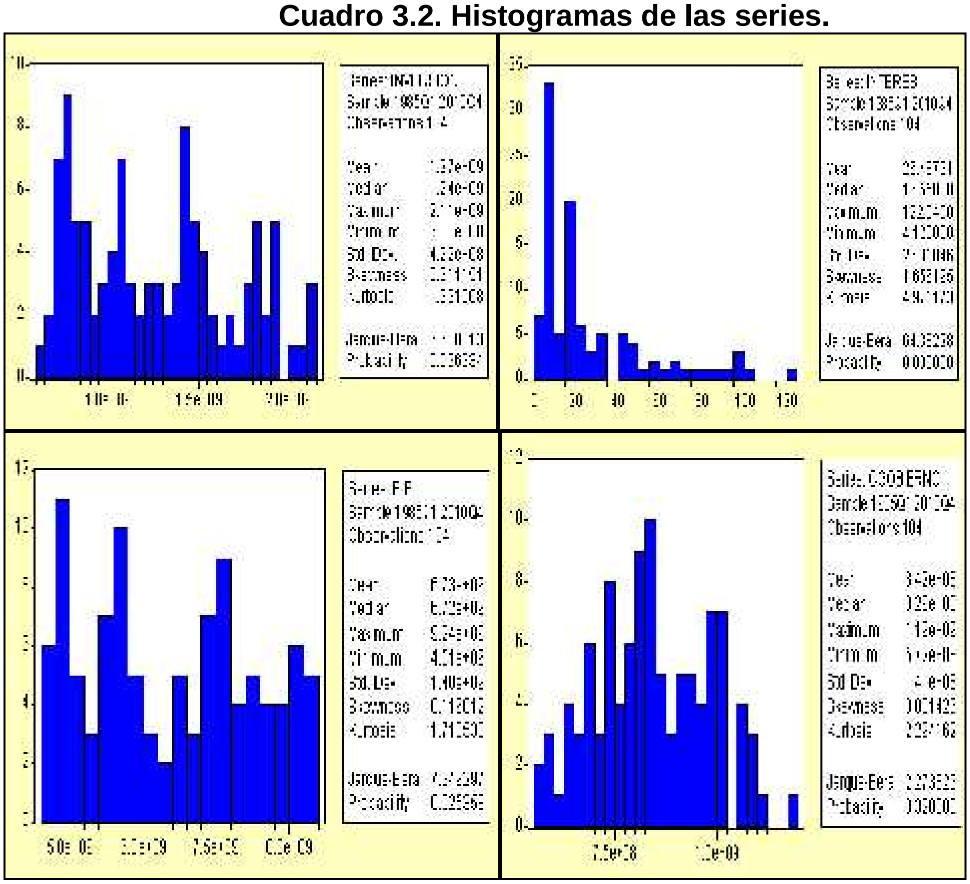
Del Cuadro 3.2 observamos que todas las variables utilizadas en el modelo (Inversión, Interés, Producto Interno Bruto y el Consumo del Gobierno) no se comportan como una normal ya que la media no es igual a la mediana, por ello no son simétricas. En cuanto al dato de asimetría todas las variables tienen asimetría positiva y cargada a la derecha por tener valores mayores que cero. Con la curtosis observamos que la Inversión, el PIB y el Consumo del Gobierno son platocúrticas por su valor en curtosis menor a tres y en el caso del Interés es leptocúrtica por su valor en curtosis mayor a tres. Ninguna variable se distribuye como una normal.

Para terminar este análisis de normalidad vamos a utilizar el estadístico Jarque-Bera con la siguiente prueba de hipótesis:

Ho: La serie se distribuye como una normal.
Ha: La serie no se distribuye como una normal.

Para que las variables fueran normales JB $\cong 6$, en los Cuadro 3.1 y 3.2 se muestran los datos para cada variable del estadístico Jarque-Bera. En el caso de la Inversión, el Interés y el Producto Interno Bruto se tienen valores mayores a 6 de dicha estadística, ello nos indica que la serie no se distribuye como una normal, por tanto, se rechaza la hipótesis nula. Para el caso de la variable Consumo de Gobierno que al tener la estadística Jarque-Bera un valor de 2.2738, que es menor a 6, no se rechaza la hipótesis nula, por lo que, esta variable si presenta normalidad. Se corrobora esto al analizar las probabilidades asociadas a dicho estadístico, ya que la única mayor a 0.05 para indicar normalidad es la variable de Consumo de Gobierno.

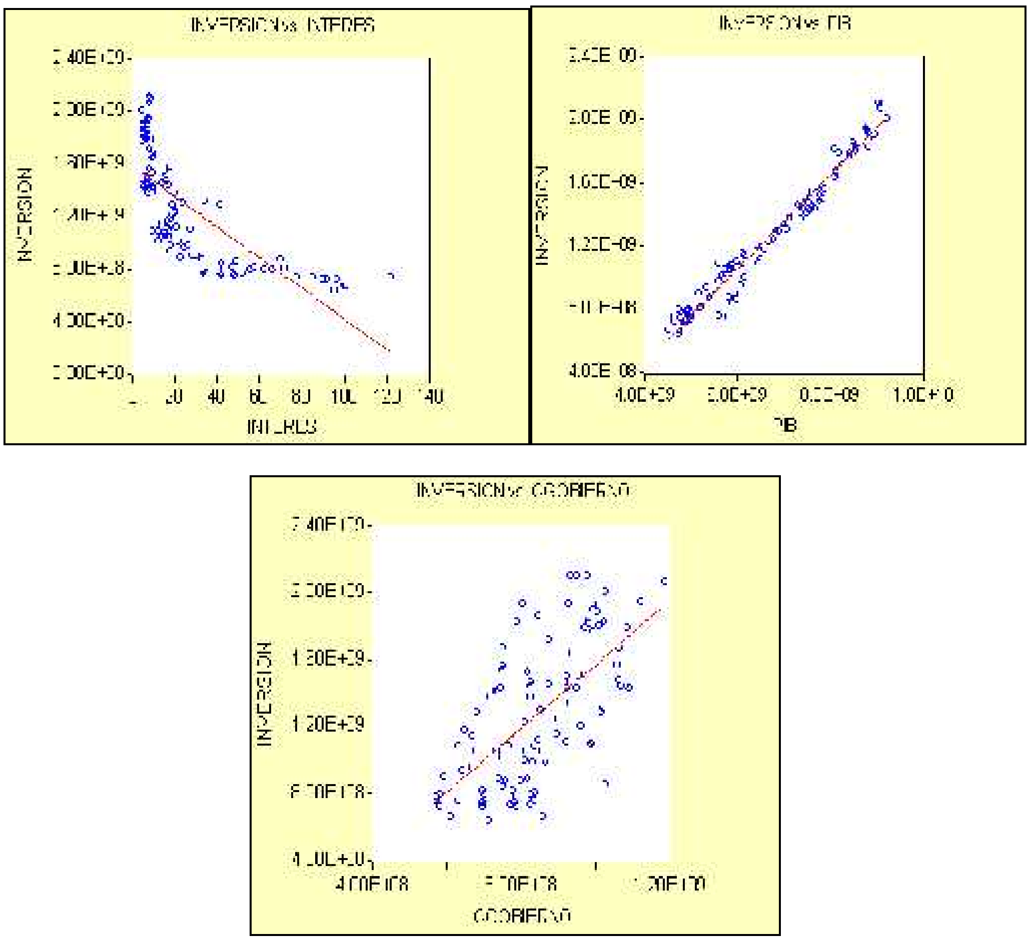
Cuadro 3.2. Histogramas de las series.



Con el histograma podemos visualizar y constatar lo antes mencionado acerca de normalidad en las series evaluando el estadístico Jarque-Bera, pero ahora con el método visual ya que la única distribución con forma de campana es la variable Consumo de Gobierno.

Ahora realizamos los diagramas de dispersión para determinar la relación funcional que siguen cada una de las variables independientes con respecto a la variable dependiente.

Gráfica 3.2. Diagramas de dispersión con regresión lineal.



Al observar la Gráfica 3.2 que contiene los diagramas de dispersión, tenemos que la relación de la Inversión con el Interés es negativa como lo habíamos explicado anteriormente, aunque su relación no es lineal al no coincidir con la línea de regresión que muestra una relación lineal perfecta. En cuanto a la Inversión con el Producto Interno Bruto tenemos que la relación si es lineal al coincidir nuestros datos con la línea de regresión mostrando una relación positiva y lineal, es decir, al incrementarse el Producto Interno Bruto se incrementará la Inversión por esta relación. Al pasar al diagrama que muestra la Inversión con relación al Consumo del Gobierno se torna complicado su análisis ya que aunque la relación se muestra positiva, los puntos observados en el diagrama son muy dispersos y no sigue la relación lineal de las variables, por ello se puede afirmar que el diagrama de dispersión no muestra ningún tipo de relación que sigan ambas variables.

A continuación procedemos a obtener los resultados de la regresión (ver Cuadro 21).

Cuadro 3.3. Resultados de la ecuación de regresión.

Dependent Variable: INVERSION
Method: Least Squares
Sample: 1985Q1 2010Q4
Included observations: 104

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.85E+08	67357784	-10.17204	0.0000
INTERES	-98578.63	422142.7	-0.233520	0.8158
PIB	0.300696	0.009668	31.10074	0.0000
CGOBIERNO	0.077947	0.071206	1.094664	0.2763

R-squared	0.968370	Mean dependent var	1.27E+09
Adjusted R-squared	0.967421	S.D. dependent var	4.22E+08
S.E. of regression	76149594	Akaike info criterion	39.17200
Sum squared resid	5.80E+17	Schwarz criterion	39.27371
Log likelihood	-2032.944	F-statistic	1020.512
Durbin-Watson stat	0.308269	Prob(F-statistic)	0.000000

La ecuación de regresión como lo muestra el Cuadro 3.3 es la siguiente:

$$Inversión = -685166181.8 - 98578.6{*}Interes + 0.3{*}PIB - 0.07{*}\ C\ Gobierno$$

En primera instancia podemos observar que el coeficiente â es negativo, esto no concuerda con la teoría planteada en el ejercicio 3 de regresión simple. Ahí se afirmaba que â debe de ser igual o mayor a cero, dado que, cuando no influyen las variables independientes incluidas en el modelo, la Inversión autónoma para que la economía siga funcionando debe ser positiva. Este resultado incongruente puede ser debido a una mala especificación del modelo en cuanto a la selección de la forma funcional, ya que esta es lineal y como lo hemos observado en los diagramas de dispersión, la relación de la Inversión con la Tasa de Interés no es perfectamente lineal y con el Consumo del Gobierno no tiene tendencia.

Pasando ahora a los estimadores de las variables independientes vemos que todos cumplen con la teoría ya que el estimador del Interés ($\hat{b}_1$) es negativo, afirmándose que si la variable Interés se incrementa nuestra variable dependiente va a disminuir y viceversa. En cuanto al estimador del PIB ($\hat{b}_2$), tiene signo positivo mostrando que ante incrementos en el Producto Interno Bruto la Inversión va a incrementarse por la relación positiva entre estas variables. Por último tenemos el estimador del Consumo Gubernamental ($\hat{b}_3$) el cual también cumple la teoría neoclásica que dice que la Inversión con el Consumo del Gobierno es negativa dado que al incrementarse la participación del Gobierno se reduce la participación del sector privado y por lo tanto la Inversión.

Analizando las cifras resultantes de la regresión tenemos que el coeficiente de Inversión Autónoma, â, es de -685166181.8 lo que nos dice que si no influyen las variables independientes incluidas en el modelo como lo es el Interés, el Producto Interno Bruto y el Consumo del Gobierno la Inversión será negativa, lo que no concuerda con la teoría, ya que es ilógico que la inversión sea negativa.

Con los coeficientes de las variables independientes se muestra el grado de afección de éstas en la Inversión. Tenemos así que si la Tasa de Interés se incrementa en una unidad porcentual, la Inversión se verá disminuida en 98578.6 miles de pesos constantes del 2003, permaneciendo las demás variables independientes constantes y viceversa. En el caso del Producto Interno Bruto si este se incrementa en mil pesos constantes de 2003, la Inversión aumentará 0.3006 miles de pesos constantes de 2003, lo cual tiene sentido teórico ya que al incrementarse los ingresos se incrementa la inversión y viceversa. Por último se tiene que si se incrementa el Consumo del Gobierno en mil pesos constantes de 2003 la inversión disminuirá 0.0779 miles de pesos constantes de 2003 por el sentido negativo de las variables y viceversa.

Por lo que respecta a la significación estadística de los estimadores, planteamos la siguiente prueba de hipótesis para los estimadores en general:

Ho: $_1 = 0$, por tanto, la variable independiente no explica a la Inversión.

Ha: $_1 \neq 0$, por tanto, la variable independiente explica a la Inversión.

Los criterios, con un nivel de significancia al 5%, para rechazar o no la Ho con el p-valué son:

1. Si la probabilidad asociada al estadístico “t” es menor o igual a 0.05 se rechaza la Ho.
2. Si la probabilidad asociada al estadístico “t” es mayor a 0.05 no se rechaza la Ho.

Por tanto, al ser la probabilidad asociada al estadístico “t” del Interés mayor a 0.05, con un valor de 0.8158 no se rechaza la Ho y por tanto el Interés no explica a la Inversión con un 95% de probabilidad de que así sea. Dicha conclusión puede ser atribuible a la mala especificación del modelo, es decir, a la incorrecta selección de la forma funcional entre estas dos variables o a la posible violación a alguno de los supuestos del modelo de estimación.

Para el estimador del Producto Interno Bruto la probabilidad asociada al estadístico t de dicho estimador es menor que 0.05 por lo que se rechaza la hipótesis nula (Ho) mostrando que el PIB explica a la Inversión al 95% de probabilidad de que así sea.

Para el caso del estimador del Consumo del Gobierno tenemos que la probabilidad asociada al estadístico t es de 0.2763, lo que es mayor a 0.05, por lo que no se rechaza la hipótesis nula afirmando que el Consumo del Gobierno no explica a la Inversión al 95% de probabilidad de que así sea.

En cuanto a la prueba de hipótesis para la inversión autónoma, tenemos:

Ho: α = 0, por tanto, su valor no tiene o no es de relevancia económica.
Ha: α > 0, por tanto, su valor si es de relevancia económica.

Así, al ser la probabilidad asociada a la estadística “t” del estimador $\hat{a}$ menor a 0.05, se rechaza la Hipótesis nula, y por tanto, el valor de la inversión autónoma es de relevancia económica. Aunque la prueba nos afirma que es de relevancia económica el valor es negativo, no mostrando congruencia teórica ni relevancia para nuestro modelo como lo habíamos planteado anteriormente. Ello puede ser atribuible a una mala especificación del modelo.

En este ejercicio al hablarse de regresión múltiple se incluye una prueba de significación global o también llamada prueba “F” ya que es obtenida a partir de la estadística “F”. Para ello se plantea una prueba de hipótesis conjunta:

Ho: ${}_1 = {}_2 = {}_3 = 0$, por lo que, el Interés, el PIB y el Consumo del Gobierno no explican a la Inversión.
Ha: ${}_1 \neq {}_2 \neq {}_3 \neq 0$, por lo que, el Interés, el PIB y el Consumo del Gobierno explican a la Inversión.

La estadística “F” viene dada por:

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

Así, para nuestro caso, k=4, n=104 y R^2=0.968370, por lo que:

$$F = \frac{0.96837/(4-1)}{(1-0.96837)/(104-4)}$$

$$F = \frac{0.96837/3}{0.03163/100} = \frac{100(0.96837)}{3(0.03163)}$$

$$F = \frac{96.837}{0.09489} = 1020.5185$$

El valor obtenido o calculado de “F” se contrasta con la “F_α”, también conocida como estadística “F” teórica o de tablas.
Así, la estadística “F” teórica es:

$$F_{3,100} = 2.6955$$

La estadística “F” calculada se contrasta con el valor de la estadística “F” teórica, para ello se establecen los siguientes criterios de decisión:

1. Si “F” calculada es menor que la “F” teórica no se rechaza la Ho.
2. Si “F” calculada es mayor o igual a “F” teórica se rechaza la Ho.

Como “F” calculada es mayor que la “F” teórica, 1020.5185 > 2.6955, se rechaza la Ho, afirmando que el Interés, el Producto Interno Bruto y el Consumo del Gobierno si explican a la Inversión.

Este resultado puede ser corroborado mediante la probabilidad asociada a la estadística “F” que el programa proporciona (p-value). Los criterios de decisión son los siguientes:

1. Si la probabilidad asociada a la estadística “F” es menor o igual a 0.05 se rechaza la Ho.
2. Si la probabilidad asociada a la estadística “F” es mayor a 0.05 no se rechaza la Ho.

Por lo que, al ser la probabilidad asociada a la estadística “F” menor a 0.05, con un valor de 0.0000 que se observa en el Cuadro 21, se rechaza la Ho y se corrobora la conclusión de que las variables independientes incluidas en el modelo explican a la variable dependiente, en este caso la Inversión.

Dicha conclusión, a partir de la prueba de hipótesis conjunta del modelo, no coincide con el resultado de las pruebas de significación individual, ya que con éstas últimas se concluía que el Interés y el Consumo del Gobierno no tienen relación con la variable dependiente, cabe recordar que estas son pruebas individuales, la prueba de significación global “F” analiza a las variables en conjunto afirmando que las variables incluidas en el modelo si explican a la variable dependiente. Esta incongruencia puede ser atribuible, en principio, al incorrecto uso de formas funcionales lineales entre variables independientes con la dependiente o al incumplimiento de alguno o algunos de los supuestos detrás del método de estimación.

En cuanto a la prueba de bondad de ajuste tenemos que: R^2 = 0.968370, lo que significa que el 96.8370% de las variaciones de la Inversión son explicadas por las variables independiente incluidas en el modelo. Mientras que $\overline{R}^2$=0.967421, lo que significa que el 96.7421% de los cambios en la Inversión se explican por cambios en las variables incluidas en el modelo y una variable adicional.

Ahora procedemos a analizar los coeficientes de correlación parciales mediante la matriz de correlaciones que se muestran en el Cuadro 3.4.

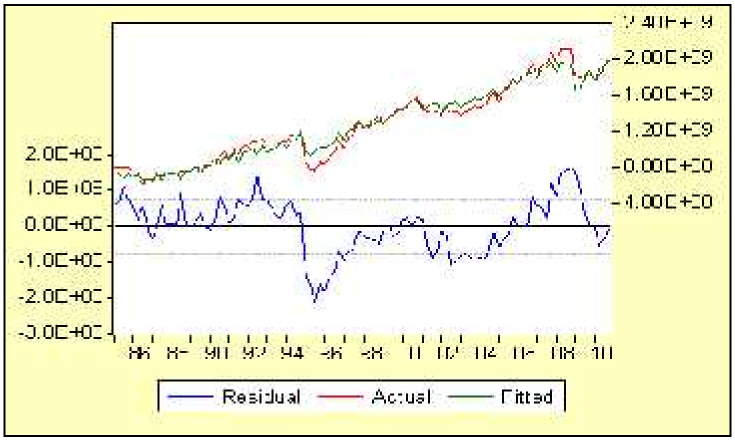
Cuadro 3.4. Matriz de correlaciones.

	INVERSION	INTERES	PIB	CGOBIERNO
INVERSION	1.000000	-0.737698	0.983846	0.634391
INTERES	-0.737698	1.000000	-0.745650	-0.441820
PIB	0.983846	-0.745650	1.000000	0.660068
CGOBIERNO	0.634391	-0.441820	0.660068	1.000000

En cuanto al coeficiente de correlación entre la Inversión y el Interés observamos que este asciende a -0.7376, por lo que puede concluirse que su relación es aceptable ya que es mayor que 0.5 con una relación inversa entre ambas variables. La correlación entre la Inversión y el Producto Interno Bruto es de 0.983846 lo que indica una alta correlación ya que es cercano a uno, además de que su relación es positiva. Por lo que respecta al valor del coeficiente de correlación entre la variable explicada y el Consumo del Gobierno éste asciende a 0.634391 que indica correlación aceptable entre estas dos variables. Sin embargo, el signo del coeficiente no coincide con el obtenido en la ecuación de regresión. Ello puede ser debido al uso incorrecto de relaciones lineales entre variables y a la posible violación de los supuestos del modelo de estimación.

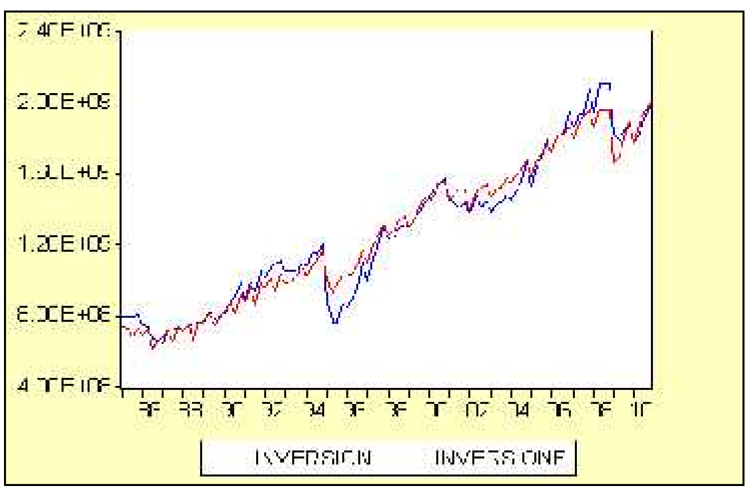
Finalmente, por lo que respecta a los valores estimados de la Inversión a través de la ecuación de regresión ($\hat{Y}$), estos se obtienen al sustituir los valores que asumen las variables independientes, como lo es en este caso el Interés, el Producto Interno Bruto y el Consumo del Gobierno en dicha ecuación. A continuación se muestra gráficamente la estimación de la Inversión mediante la ecuación de regresión obtenida, su comparación con los datos reales de la misma variable y sus errores (ver Gráfica 3.3).

Gráfica 3.3. Representación gráfica de los valores reales y estimados de la variable dependiente y los errores.



En la Gráfica 3.3 observamos como los errores son notables entre 1994 y 1998 dadas las situaciones que analizábamos anteriormente, ya que en tiempos de crisis, las variables pueden comportarse de manera distinta o atípica. En la comparación de la Inversión real y la estimada en la Gráfica 3.4 se observa que los datos reales y los estimados son muy semejantes, con la diferencia de que en los periodos de crisis, que coinciden con bajas notables en la Inversión, la serie estimada con la ecuación de regresión tiende a hacer menos abruptas dichas caídas, con la excepción del periodo 2008-2009.

Gráfica 3.4. Gráfica del consumo y el consumo estimado.



Con esta exposición termina la presentación teórica y práctica del MLG, esperando que el lector haya captado los aspectos básicos de sus conceptos y métodos, así como la forma correcta en que se hace el análisis e interpretación económica y estadístico del MLG.

V.2.9.- Reactivos para reafirmar los conocimientos.

A.- Planteamiento de preguntas y respuestas

Si el análisis de regresión múltiple es con dos variables independientes: a).- indique el significado de α, β_1, y β_2

Respuestas:

α es el término constante u ordenada al origen de la ecuación de regresión y proporciona el valor estimado de Y cuando $X_1 = X_2 = 0$;

El valor del parámetro ,β_1 expresa el cambio ocasionado en Y por cada unidad que cambia X_1 siempre que X_2 se mantenga constante. En otras palabras, β_1 es la pendiente de X_1 que por eso también se le denomina *coeficiente de regresión parcial* porque corresponde a la derivada parcial de Y con respecto a X_1.

Por consiguiente, con ese mismo razonamiento se dice que el valor del parámetro β_2 representa el cambio en Y ocasionado por el cambio en una unidad de X_2, siempre que X_1 se mantenga constante. Así, β_2 es la pendiente de X_2, misma que también es un coeficiente de regresión parcial como lo es β_1 dado que corresponde a la derivada parcial de Y con respecto a X_2.

b).- ¿Cuándo se dice que esos tres parámetros son significativos estadísticamente?

Respuestas:

Individualmente son significativos estadísticamente cuando su **tα** teórica o de tablas es menor que su **t** real o calculada; lo son *globalmente,* cuando su **Fα** teórica o de tablas es menor que su **F** real o calculada.

En los dos casos se toma la decisión de aceptar o rechazar la Ho, i.e., de indicar que dichos coeficientes son o no significativamente diferentes de cero con cierto nivel de significación (denotado por α, que no es la misma α de la ecuación de regresión múltiple) y ciertos grados de libertad previamente especificados.

La interpretación estadística, *individual*, es que se rechaza la hipótesis nula de que cada una de tres son iguales a cero y por consiguiente, que se acepta la hipótesis alternativa de que $\alpha \neq 0$, $\beta_1 \neq 0$ y $\beta_2 \neq 0$, respectivamente. En *grupo,* que se rechaza la hipótesis nula de que $\alpha = \beta_1 = \beta_2 = 0$,; ello significa que se acepta la hipótesis alternativa de que todas ellas son distintas de cero y que simultáneamente, R^2 es significativamente diferente de cero: , $R^2 \neq 0$.

c).- En términos económicos, ¿qué significa que α, $\beta 1$, y β_2 sean significativos estadísticamente?

Respuesta:

La interpretación económica es que se verifica la hipótesis de que X_1 y X_2 si explican a Y

d).- ¿Cuándo es conveniente verificar la significación estadística de α?

Respuesta:

Es conveniente cuando la teoría económica en estudio le asigna un papel relevante. Cuando no es así, por lo general suele omitirse la prueba de hipótesis sobre su significación estadística.

Examen de autoevaluación 1.

Obtenga la Ecuación de Regresión y el Coeficiente de Correlación Múltiples

Suponga que la teoría económica establece que la rentabilidad de la inversión (Y) depende de la tasa de interés (X) y de la inflación (Z), tal que:
$\hat{Y} = f(X, Z)$

Tomando como referencia los datos para Y_i, X_i, Z_i y si tenemos.

Año	***Y_i***	***X_i***	***Z_i***								
1	7	9	1								
2	8	8	2								
3	7	9	3								
4	8	7	2								

Preguntas:
1.- Obtenga la ecuación de regresión: $\hat{Y} = a + bX_i + cZ_i$
2.- Encuentre el valor para cada $\hat{y}_i$ donde i=1,2,3,4:
3.- Encuentre el valor del error residual
4.- Pruebe la hipótesis nula de que la rentabilidad de la inversión depende de la tasa de interés y de inflación, con un nivel de significación del 5% e interprete los resultados.
5.-Calcule R, $R^{2,}$ $\overline{R}^2$ y junto con t, F, y con el error estándar de estimación interprete sus resultados.

Comentarios: Al resolverse el examen el lector hallará que este ejemplo es muy interesante porque sus resultados estadísticos al ser incongruentes revelan entre otras cosas: a). La importancia de la especificación correcta y detallada de la teoría económica; b). la importancia que tiene el tamaño de la muestra y c).-.La conveniencia de revisar si se cumplen los supustos téoricos que fundamentan el uso del método de MCO para calcular los estimadores muestrales de los parámetros poblacionales.

Examen de auto evaluación 2:

Tema: Marco teórico de la expresión matemática de una teoría económica y de su verificación estadística.

I.-Conteste con una “X” en **SI** cuando la afirmación sea verdadera y también con una “X” en NO cuando la afirmación sea falsa:

1.- Una variable cualitativa no se puede expresar cuantitativamente: Si___; No____.

2.-Una variable dicotómica puede tomar más de dos valores: Si_____; No________

3.-Para predecir la variable dependiente es necesario que antes se conozcan los valores proyectados de las variables explicativas en el tiempo: SI:_______; NO:___________

4.- Cuando se verifica la hipótesis de la relación entre la variable dependiente con las independientes, se espera que el valor del coeficientes de correlación múltiple no sea semejante a los valores de cada uno de los coeficientes de correlación parcial: SI: __; NO__.

5.- Cuando se verifica con la estadística **F** que las variables independientes sí explican a la variable dependiente, teóricamente se debe de verificar con la prueba ***t*** de Student que cada una de las variables independientes también explican a la dependiente: SI___; NO____.

6.- Al correr la regresión, cuando los signos de los coeficientes obtenidos de las variables regresoras no coinciden con la concepción teórica de su relación con la variable regresada, de inmediato se debe de cambiar de variable regresora : SI____; NO______:

7.- Al probar una hipótesis nula con la ***t*** de Student, si la ***t*** real o calculada es mayor que la ***t*** teórica o de tablas, se acepta dicha hipótesis nula: SI____;NO_____.

8.- Cuando F real o calculada es menor que F teórica o de tablas rechazamos la hipótesis nula: SI_____; NO____.
:
9.- en una regresión múltiple si al probar una hipótesis nula con ciertos grados de libertad y nivel de significación, cuando los valores de t no son estadísticamente significativos pero F si lo es, se dice que posiblemente se violó algún de los supuestos del modelo de estimación: SI____; NO______.

10.-Conocido el valor del coeficiente de la variable explicativa X, su media aritmética como la de la variable explicada Y, la elasticidad se calcula así: $E = \hat{b}\frac{\bar{Y}}{\bar{X}}$: SI__;NO___.

Observaciones: Cada una de las respuestas correctas vale 10 puntos en escala de 0 a 100. Al ser un examen sobre conocimientos básicos,no se puede usar la calculadora, ni las tablas estadísticas ni la bibliografía correspondiente a cada tema.

VI.- Las violaciones a los supuestos del metodo de MCO.

Como se recordará, este método se usa para obtener estimadores contenidos en la ecuación de regresión, considerando que se cumplen ciertos supuestos que sustentan sus **propiedades** de ser: a).- insesgados; b).- eficientes, c).- suficientes y d).- consistentes, principalmente.

Dicho método se apoya, entre otros, en las siguientes hipótesis o supuestos:

1.- la varianza de las Ui es constante y por ello se dice que hay **homoscedasticidad**, que viene del griego: homos (igual) y cedastitis (dispersión) entre los miembros de la serie estadística, razón por la cual tienen la varianza mínima, que a su vez corresponde a los estimadores que hemos dado en llamar eficientes.

2.- No existe **autocorrelación** entre las perturbaciones o residuos, Ui, y

3.- No existe **multicolinialidad** entre las variables explicativas de la ecuación de regresión.

4.- El modelo de regresión está perfectamente **especificado**, de manera que no existe ningún

Sesgo de especificación (Gujarati, 1991:210).

Cuando se cumplen estos y otros supuestos se tiene una buena inferencia estadística y se está en condiciones de hacer una adecuada estimación y mejores pruebas de hipótesis, y por consiguiente, una toma de decisiones adecuada.

¿Pero qué sucede cuando se violan estos supuestos?

Definitivamente se pierde calidad en los *estimadores* y disminuye el rigor técnico con que se maneja la información ya que dejan de ser insesgados, eficientes, consistentes y suficientes, afectando por medio de sus errores estándar la estimación que se hace con la ecuación de regresión y orillando al investigador a la toma equivocada de decisiones cuando se hacen pruebas de hipótesis porque las t's y las F´s cambian de valor, en la forma que se indica a continuación:

VI.1 Heteroscedasticidad.

Uno de los principales análisis que se realizan sobre la violación de los supuestos en que se basa el método de MCO para determinar el valor y por consiguiente la calidad de los estimadores, se refiere a la verificación, Ho, de si las perturbaciones Ui de la función de regresión poblacional, son homoscedásticas, i.e., que todas tienen la misma varianza: que ésta se mantiene a lo largo de las observaciones; en otras palabras, es conveniente señalar que hasta el momento hemos establecido el supuesto de homoscedasticidad, es decir, que las distorsiones o errores Ui de la ecuación de regresión tienen la misma varianza.

Ahora bien, cuando dichos errores no observan una misma varianza se acepta la Ha y se dice que hay heteroscedasticidad o que las Ui son heteroscedásticas. En otras palabras las Ui no tienen una varianza constante, que es lo mismo que decir que la varianza del error no es constante para todas las observaciones de la serie histórica a partir de la que se determinó la ecuación de regresión.

Precisamente con un enfoque de datos correspondientes a series de tiempo es que César Pérez (2007: 134) comenta que la Ho: Cualquier momento t de la Ui tiene media cero y varianza independiente de t, así como que Cov (Ui,Uj)=0 para toda i y todo j momentos temporales distintos entre sí. Por eso es que a Ho le denomina hipótesis nula de *homoscedasticidad condicional*, i.e., no hay variaciones estadísticas significativas, misma que verifica usando las técnicas ARCH y GARCH.

VI.1.1.- ¿Por qué surge? ¿Cuáles son sus orígenes?

Con Frecuencia aparece en datos de corte transversal (Carrascal et al, 2001) porque las unidades de muestreo no tienen un comportamiento homogéneo, al igual que también surge en series de tiempo no estacionarias; también se origina en datos cuyos valores se han obtenido agregando o promediado (medias móviles) datos individuales

VI.1.1.1.-¿Qué efecto o consecuencia trae la heteroscedasticidad?

- Las estimaciones $\hat{a}$ y $\hat{b}$ de mínimos cuadrados son insesgados pero no consistentes ni eficientes, es decir, no poseen varianza mínima, algunos datos tienen una varianza más grande; además el valor del estimador no tiende al del parámetro a medida que crece el tamaño de la muestra, por lo que se le califica de inconsistente.
- Las varianzas estimadas var ($\hat{a}$), var ($\hat{b}$) no son insesgadas. Al ser sesgados los estimadores de las varianzas, invalidan las pruebas de significación sobre las hipótesis que se establezcan.

VI.1.2.- ¿Cómo se detecta?

Señala Gujarati (1990:258) que no es fácil detectarla, que “**no existen reglas fijas y seguras para detectarla, sino sólo unas cuantas reglas generales”;** por ello se han creado algunos métodos informales y de aproximación para detectar la presencia de heteroscedasticidad, los cuales generalmente examinan los **residuos** obtenidos de la aplicación de MCO **para identificar en ellos patrones sistemáticos.** De manera que si las U_i presentan dichos patrones, se está en condiciones de recomendar procedimientos para transformar los datos originales bajo estudio, tal que con la nueva ecuación de regresión con datos ya transformados las U_i ya observen una misma varianza, es decir varianza constante.

Lo anterior en palabras de Carrascal (2001:227): “no existen reglas fijas para saber si en un modelo existe heteroscedasticidad; pues en todos los contrastes estadísticos se plantea la hipótesis nula de homoscedasticidad. Además, dado que las perturbaciones aleatorias no son observables, las formas de detección se basan en los *errores* de la estimación por mínimos cuadrados ordinarios. En concreto, la mayor parte de los contrastes van a utilizar el cuadrado de dichos errores como indicativo de la varianza de cada perturbación o el valor absoluto de dicho error para aproximar la desviación típica.”

Derivado de lo anterior podemos decir que en general se pueden realizar cualquiera de las siguientes pruebas:

1. Método gráfico
2. Ramsey: RESET
3. Glejser
4. Breusch y Pagan
5. White
6. Goldfeld y Quant
7. Razón de verosimilitudes

Al respecto sobre el **método gráfico**, G.S Maddala (1996: 229), pone un ejemplo sencillo pero ilustrativo a través del cual se identifica la heteroscedasticidad con el método gráfico, **dado que son datos de corte transversal.**
El autor hace función el consumo (y) del ingreso (x). Para ello presenta la información de 20 familias, misma que aparece en la siguiente tabla, cuyo ingreso y consumo se expresa en miles de dólares.

FAMILIA	Y	Y_c	X	$Y-Y_c=U_i$ RESIDUO
1	19.9	20.9019921708	22.3	-1.00
2	31.2	29.8952390494	32.3	1.30
3	31.8	33.7623352071	36.6	-1.96
4	12.1	11.7288803547	12.1	0.37
5	40.7	38.8884859279	42.3	1.81
6	6.1	6.42286469639	6.2	-0.32
7	38.6	41.0468651788	44.7	-2.45
8	25.5	24.3194259847	26.1	1.18
9	10.3	10.1100959166	10.3	0.19
10	38.8	36.9999040834	40.2	1.80
11	8.0	8.13158160331	8.1	-0.13
12	33.1	31.8737533626	34.5	1.23

13	33.5	35.0213897701	38.0	-1.52
14	13.1	13.5275297304	14.1	-0.43
15	14.8	15.5959765125	16.4	-0.80
16	21.6	22.520776609	24.1	-0.92
17	29.3	27.9167247361	30.1	1.38
18	25.0	26.297940298	28.3	-1.30
19	17.9	17.2147609506	18.2	0.69
20	19.8	18.9234778576	20.1	0.88

Corrió la regresión y obtuvo:

$Y = 0.847 + 0.899X \qquad R^2 = 0.986$

$(0.703) \quad (0.0253) \qquad RSS = 31.074$

Dicha ecuación proviene de la siguiente tabla obtenida con Eviews:

Dependent Variable: Y
Method: Least Squares
Sample: 1901 1920
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.847052	0.703355	1.204302	0.2441
X	0.899325	0.025309	35.53360	0.0000

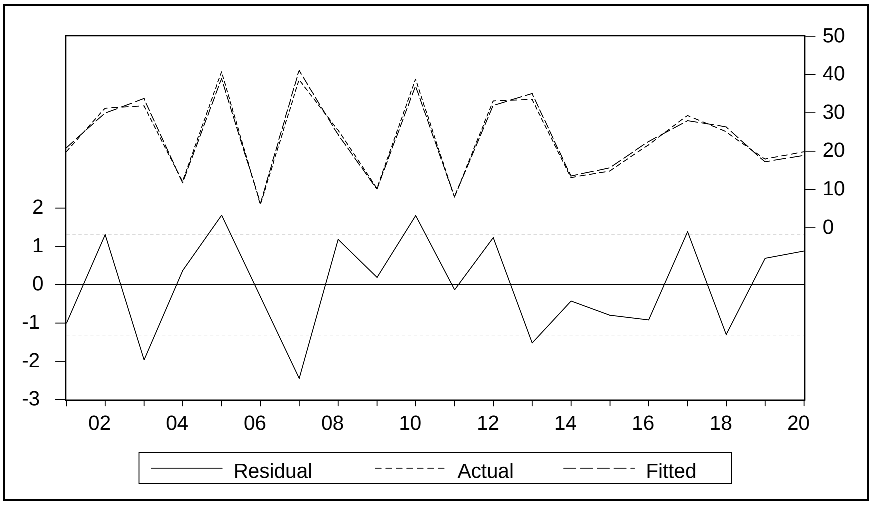
R-squared	0.985944	Mean dependent var	23.55500
Adjusted R-squared	0.985164	S.D. dependent var	10.78691
S.E. of regression	1.313895	Akaike info criterion	3.478509
Sum squared resid	31.07377	Schwarz criterion	3.578082
Log likelihood	-32.78509	F-statistic	1262.637
Durbin-Watson stat	2.582686	Prob(F-statistic)	0.000000

Con esos datos procedemos a identificar la heteroscedasticidad:

VI.1.2.1.- Métodos Gráficos.

Para calcular Ui: en el Cuadro de la ecuación, está la palabra *view,* ahí pulsamos el cursor y aparece, entre otros, *actual fitted residuals*, pedimos *actual fitted table,* oprimimos el lado izquierdo del ratón y parecen los valores originales de Y, los de cada una de las Y´s calculadas con la ecuación de regresión anterior y Ui= Yi-Yc donde i=1,2,3,…….18,19,20

Estando en la pantalla de este archivo, vamos a vi*ew,* ahí pedimos *actual fitted residuals,* luego, *actual fitted graph,* decimos *ok,* y aparece la gráfica de residuos siguiente

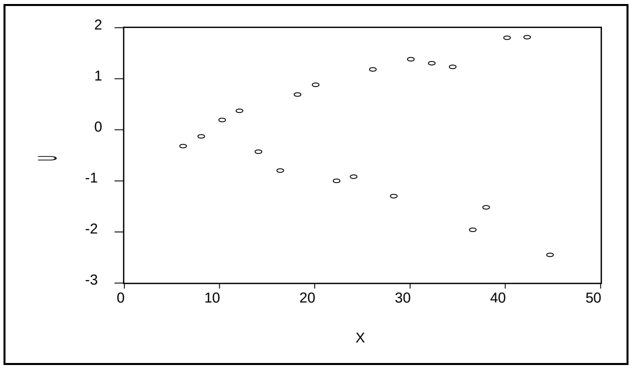


Ahora lo anterior también se puede ver en la siguiente forma: vamos a Quick/ Estimate Equation; U- C X,/ ok y sale la ecuación. Ahora vamos a View Actual, Fitted, Residual/ Actual Fitted Table y sale:

<table>
<tr><th>Obs</th><th>Actual</th><th>Fitted</th><th>Residual</th><th>Residual Plot</th></tr>
<tr><td>1</td><td>19.9</td><td>20.9019921708</td><td>-1.00199217083</td><td>| .* | . |</td></tr>
<tr><td>2</td><td>31.2</td><td>29.8952390494</td><td>1.30476095063</td><td>| . | * |</td></tr>
<tr><td>3</td><td>31.8</td><td>33.7623352071</td><td>-1.96233520714</td><td>| * . | . |</td></tr>
<tr><td>4</td><td>12.1</td><td>11.7288803547</td><td>0.371119645276</td><td>| . | * . |</td></tr>
<tr><td>5</td><td>40.7</td><td>38.8884859279</td><td>1.8115140721</td><td>| . | . * |</td></tr>
<tr><td>6</td><td>6.1</td><td>6.42286469639</td><td>-0.322864696388</td><td>| . * | . |</td></tr>
<tr><td>7</td><td>38.6</td><td>41.0468651788</td><td>-2.44686517875</td><td>|* . | . |</td></tr>
<tr><td>8</td><td>25.5</td><td>24.3194259847</td><td>1.18057401532</td><td>| . | * |</td></tr>
<tr><td>9</td><td>10.3</td><td>10.1100959166</td><td>0.189904083412</td><td>| . |* . |</td></tr>
<tr><td>10</td><td>38.8</td><td>36.9999040834</td><td>1.80009591659</td><td>| . | . * |</td></tr>
<tr><td>11</td><td>8</td><td>8.13158160331</td><td>-0.13158160331</td><td>| . *| . |</td></tr>
<tr><td>12</td><td>33.1</td><td>31.8737533626</td><td>1.22624663735</td><td>| . | * |</td></tr>
<tr><td>13</td><td>33.5</td><td>35.0213897701</td><td>-1.52138977013</td><td>| *. | . |</td></tr>
<tr><td>14</td><td>13.1</td><td>13.5275297304</td><td>-0.427529730432</td><td>| . * | . |</td></tr>
<tr><td>15</td><td>14.8</td><td>15.5959765125</td><td>-0.795976512495</td><td>| . * | . |</td></tr>
<tr><td>16</td><td>21.6</td><td>22.520776609</td><td>-0.920776608968</td><td>| .* | . |</td></tr>
<tr><td>17</td><td>29.3</td><td>27.9167247361</td><td>1.38327526391</td><td>| . | .* |</td></tr>
<tr><td>18</td><td>25</td><td>26.297940298</td><td>-1.29794029795</td><td>| * | . |</td></tr>
<tr><td>19</td><td>17.9</td><td>17.2147609506</td><td>0.685239049368</td><td>| . | * . |</td></tr>
<tr><td>20</td><td>19.8</td><td>18.9234778576</td><td>0.876522142446</td><td>| . | * . |</td></tr>
</table>

Lo anterior ahora visto en términos de dispersión de las U_i con respecto a X's:

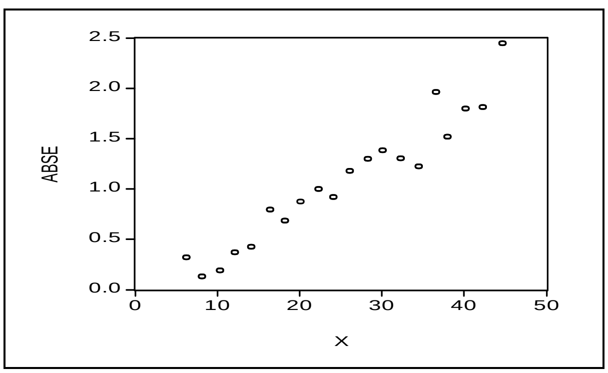
Continuando con el análisis gráfico, **ahora** sí representamos la relación X_i con U_i vamos a *Quick* ahí pedimos *Graph*, aparece la pantalla Series List con *Group01*, lo borramos y en su lugar ponemos X U, damos ok y aparece *Line Graph*: seleccionamos *Scatter Diagram*, ok aparece la siguiente gráfica.



que es la figura típica que resulta al graficar los valores de los residuos versus los valores de X, ingreso, obteniéndose el diagrama que indica o permite identificar que hay un problema de heteroscedastidad, puesto que hay una alta dispersión de U_i a medida que aumenta el valor de X, mismo que debe resolverse para recuperar la bondad estadística de los estimadores.

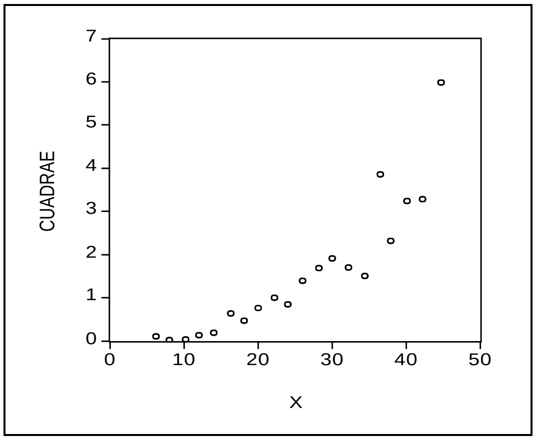
Conviene reiterar, como se estableció antes, que los datos entre paréntesis que acompañan la ecuación de regresión, corresponden a los errores estándar de los coeficientes. Así, a partir de la ecuación de regresión se calcularon los residuos en la forma ya familiar en la forma detallada antes explicada. Su análisis reveló que dichos residuos (en valores absolutos) eran más grandes a medida que crecía el valor de X, ingreso, y pequeños a medida que X tomaba valores pequeños. Esta evidencia le permitió señalar a Maddala que las varianzas de los errores no son las mismas, constantes y, por consiguiente hay heterocedasticidad, de tal manera que los estimadores â y $\hat{b}$ ya no son eficientes (pero si insesgados) y cuestionan seriamente los resultados a que se llega cuando se hacen pruebas de significación sobre las hipótesis en materia de regresión y correlación.

Ahora detectándola con los valores absolutos de "e": vamos a Quick/Generete Series/ y escribimos en pantalla: abse=@abs(resid)/ok y en el Workfile aparece la nueva variable "abse". A continuación vamos a Quick/Graph/scatter/ y escribimos: x abse/ok, apareciendo la siguiente gráfica:



Como en el caso anterior, se observa que a medida que aumentan los valores de "x" aumenta la dispersión de los residuos (abse), indicando que hay heteroscedasticidad.

De la misma manera se puede utilizar los residuos al cuadrado para probar si hay o no heteroscedasticidad; con Eviews: Quick/Generete Series/ y escribimos: cuadrae=resid^2/ok y aparece la nueva variable "cuadrae" en el Workfile. Para graficar vamos a: Quick/Graph/Scatter y escribimos: x cuadrae/ok apareciendo la siguiente gráfica:



Como en los dos casos anteriores, se confirma que hay heteroscedasticidad porque a medida que aumentan los valores de x también aumenta la dispersión de las "cuadrae".

VI.1.2.2.- ¿Cómo detectarla cuando hay más de una variable explicativa?

Existen varias opciones, ellas son:

1.- Correr una por una de las variables explicativas versus su variable explicada; o

2.- Seleccionar la variable explicativa que se suponga genera la variación; o

3.- Correr la variable Ui transformada ya sea en forma absoluta o al cuadrado versus la Yc.

VI.1.3.- Identificación numérica de la heteroscedasticidad.

Tomando como referencia los datos anteriores, se corren las regresiones y se establece la hipótesis nula de homoscedasticidad y se prueba que los coeficientes son o no significativos.

VI.1.3.1.- La prueba de Ramsey: RESET.

Se hace la regresión de $\hat{u}_t$ sobre $\hat{y}_t^2$, $\hat{y}_t^3$... y la prueba de significación de los coeficientes. Así, dado que existe una sola variable explicativa x, ésta se puede utilizar en lugar de $\hat{y}$ para identificar la heteroscedasticidad. Se hace la regresión de $\hat{u}_i$ sobre $x_i^2, x_i^3 \ldots x_i^n$. Si los resultados son:

$$\hat{u} = -0.379 + 0.236(10^{-2} x^2 - 0.549)(10^{-4} x^3) \; R^2 = 0.034$$

Si ninguno de los coeficientes tiene una relación t>1, como es el caso, se toma la decisión de aceptar la hipótesis nula, es decir, que no hay heteroscedasticidad, además que al ser $R^2 = 0.034$, pequeña, indica que no es fuerte la relación entre X, U_i, es decir, no hay heteroscedasticidad.

En opinión de César Pérez (2007:135), él sugiere que se detecte por etapas; en la primera, se deben de estimar los residuos Ui del modelo original, así como los valores correspondientes ajustados (estimados) de Yt (serie de tiempo). Luego para cada t se calculan las m primeras potencias de las potencias de las estimaciones de Yt.

VI.1.3.1.2.- El contraste o prueba de GLESJER.

Se basa en regresiones y tiene un mayor alcance que otros métodos porque pretende estimar la verdadera estructura de la heteroscedasticidad. César Pérez (2007:135) dice que para ello se deben estimar los residuos del modelo Ui por MCO y con ellos se realiza la regresión con la que se halla la variable que ha causado la heteroscedasticidad. Para mayor detalle se sugiere ver Carrascal et al (2001).

VI.1.3.1.3.- La prueba o contraste de White.

Se basa en la regresión de los errores mínimo cuadráticos al cuadrado, que son el indicativo de la varianza de las perturbaciones U_i, en relación un término constante (C), con los regresores, con sus cuadrados y sus productos, cruzados o no cruzados (Cross terms). Se entiende por “cruzados” al entrelazamiento de los valores de las variables independientes de la ecuación de regresión.

Pasos:

1.- Partimos de la tabla que contienen la ecuación de regresión y otros indicadores estadísticos: en esa tabla fijamos el cursor en view/residual test/White heteroskedasticity (cross terms) si es que los queremos, si no los queremos, usamos (no cross terms) /click. En los dos casos aparece una tabla que muestra los valores, hasta arriba, de *F y Obs*R-squared con sus respectivas* *probabilidades asociadas.* Ejemplo, suponga que Y es la variable explicada y X es la variable explicativa. Cuya ecuación de regresión es la siguiente y que se genera con Eviews en la forma antes descrita:

Y =0.847052 + 0.899325X, que en detalle es:

Dependent Variable: Y				
Method: Least Squares				
Sample: 1 20				
Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.847052	0.703355	1.204302	0.2441
X	0.899325	0.025309	35.53360	0.0000
R-squared	0.985944	Mean dependent var		23.55500
Adjusted R-squared	0.985164	S.D. dependent var		10.78691
S.E. of regression	1.313895	Akaike info criterion		3.478509
Sum squared resid	31.07377	Schwarz criterion		3.578082
Log likelihood	-32.78509	F-statistic		1262.637
Durbin-Watson stat	2.582686	Prob(F-statistic)		0.000000

2.- Para saber si hay homoscedasticidad, con E Views aplicamos el contraste de White(no cross terms) siguiendo las instrucciones de arriba y hallamos:

White Heteroskedasticity Test:				
F-statistic	61.23720	Probability		0.000000
Obs*R-squared	17.56228	Probability		0.000154
Test Equation:				
Dependent Variable: RESID^2				
Method: Least Squares				
Sample: 1 20				
Included observations: 20				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.487527	0.617660	0.789313	0.4408
X	-0.070145	0.054809	-1.279809	0.2178
X^2	0.003674	0.001061	3.463358	0.0030
R-squared	**0.878114**	Mean dependent var		1.553689
Adjusted R-squared	0.863774	S.D. dependent var		1.549642
S.E. of regression	0.571954	Akaike info criterion		1.857964
Sum squared resid	5.561227	Schwarz criterion		2.007323
Log likelihood	-15.57964	F-statistic		61.23720
Durbin-Watson stat	1.686205	Prob(F-statistic)		0.000000

Interpretación: Se observa hasta arriba en el ángulo superior derecho que las probabilidades de *F y de Obs*R-squared* son menores a 5%, por lo que en este caso decimos que se rechaza la hipótesis nula de que hay homoscedasticidad. Si dichas probabilidades hubieran sido ≥ 5%, se habría rechazado la hipótesis alterna y aceptado la hipótesis nula de que hay homoscedasticidad entre las Ui´s de las diferentes observaciones (que en este caso son 20) puesto que 5% es el valor de α, probabilidad de cometer error tipo I que el investigador se ha fijado para rechazar la hipótesis que es cierta: de que hay homoscedasticidad. En concreto, en ese caso, decimos que hay heteroscedasticidad.

En *congruencia* con la identificación de heteroscedasticidad a partir de la probabilidad de

*F y de Obs*R-squared,que es menor a 5%, ahora* el lector debe observar que si los valores de estas dos últimas son mayores a los valores de F teórica o de "tablas", $F\alpha$, entonces se acepta Ha. En este caso con α=5%, tenemos que $F\alpha$=2.08, luego como el valor de ésta es menor que los valores de *F (61.23720) y de Obs*R-squared (17.56228),* aceptamos Ha y decimos que hay heteroscedasticidad.

De manera complementaria, otro indicador de que existe o no heteroscedasticidad es el R^2 contenido en esta prueba, el cual se espera sea pequeño o nulo cuando existe homoscedasticidad; sin embargo, en este caso se observa que su valor es alto: 0.878114, lo cual indica que hay heteroscedasticidad: No tienen la misma varianza las U_i. Esta interpretación proviene del razonamiento que si las U_i tuvieran la misma varianza, es decir que hubiera homoscedasticidad, en ese caso las variables incluidas en la regresión como explicativas en la regresión auxiliar no deberían tener ningún poder explicativo sobre las U_i al cuadrado, de manera que R^2 debería tener valores cercanos a 0 y no a 1.

VI.1.3.1.4.- La prueba de Goldfeld y Quandt.

Se establecen:

Ho: $\sigma i^2=\sigma^2$, indicando que hay homoscedasticidad ; y

Ha: $\sigma_i^2 \neq \sigma^2$, indicando que hay heteroscedasticidad

Cuando las muestras no son grandes, se recomienda utilizar esta prueba.

Para probar la hipótesis nula se realizan los siguientes pasos:

En este caso *los errores obtenidos* en el primer ejercicio de las veinte familias cuyo ingreso (X) y consumo (Y) son conocidos, se clasifican en orden creciente o decreciente en dos grupos; el primero comprende los 10 valores pequeños de Ui obtenidos con respecto a x; el segundo, los valores más grandes de Ui. Enseguida se corre la regresión para cada uno de los dos grupos y, con estos datos, se hace la prueba F, mediante la cual se contrasta la hipótesis nula de la igualdad de las varianzas del error.

Para hacer más firme la toma de decisiones para aceptar o rechazar la hipótesis de homocedasticidad, Salvatore (1993:152) y Gujarati (1991:266) recomiendan sacar o quitar algunos datos centrales de la distribución con objeto de "acentuar la diferencia entre el grupo de varianza pequeña y el grupo de varianza grande". Sin embargo, en este caso no omitiremos ningún dato porque como dice Gujarati mismo: “la habilidad de la prueba de Goldfeld-Quant para llevar a cabo lo anterior en forma exitosa depende de la manera como se escoja c”, que es el número de datos a omitir. La decisión de eliminar datos es algo subjetivo; sin embargo, se aconseja cuidar que el número de observaciones en cada submuestra sea mayor al número de parámetros a estimar.

Así, tenemos que obtener dos grupos de datos: el primero, con los residuos pequeños, el segundo, con los residuos grandes. Para ello usando Eviews: en workfile/process/ vamos a clasif Sort Series/ escribimos X, *“ascending*”/ok. Igual hacemos para Y y se obtienen los siguientes valores de X e Y: Sus valores **aparecen en orden ascendente**, ahí luego*, sample*, doble clic, 1 10 *Estimate Equation name*: *Group01*; igual hacemos para *Group02*, donde *sample*: 11 20,

Primer Grupo				**SegundoGrupo**			
Número de observación	Y_1	Valor de X_1	Residuo u_i	Número de observación	Y_2	Valor de X_2	Residuo u_i
6	6.1	6.2	-0.32	8	25.5	26.1	1.18
11	8.0	8.1	-0.13	18	25.0	28.3	-1.30
9	10.3	10.3	0.19	17	29.3	30.1	1.38
4	12.1	12.1	0.37	2	31.2	32.3	1.30
14	13.1	14.1	-0.43	12	33.1	34.5	1.23
15	14.8	16.4	-0.80	3	31.8	36.6	-1.96
19	17.9	18.2	0.69	13	33.5	38.0	-1.52
20	19.8	20.1	0.88	10	38.8	40.2	1.80
1	19.9	22.3	-1.00	5	40.7	42.3	1.81
16	21.6	24.1	-0.92	7	38.6	44.7	-2.45

Se estiman estas dos regresiones $\hat{Y}_1$ y $\hat{Y}_2$

$\hat{Y}_1$ = 1.0533 + 0.876x R^2 = 0.985 ; $\hat{Y}_2$ = 3.279 + 0.835x R^2 = 0.904

(0.616) (0.038) ±**= 0.689519 ;** $\pm^2 = 0.475$ (3.443) (0.096) **=1.775825 ;** $\hat{\pm}^2 = 3.154$

Con E Views el desglose estadístico de estas dos regresiones es, empezando con Y_1 , X_1 :

Dependent Variable: Y1				
Method: Least Squares				
Sample(adjusted): 1 10				
Included observations: 10 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.053316	0.616164	1.709474	0.1257
X1	0.876016	0.037939	23.09013	0.0000
R-squared	0.985217	Mean dependent var		14.36000
Adjusted R-squared	0.983369	S.D. dependent var		5.346692
S.E. of regression	**0.689519**	Akaike info criterion		2.271210
Sum squared resid	3.803487	Schwarz criterion		2.331727
Log likelihood	-9.356051	F-statistic		533.1539
Durbin-Watson stat	1.745354	Prob(F-statistic)		0.000000

Significado *S.E. of Regressión*= $\pm_{YX}$ que antes usamos; es distinto de *Std. Error* que suele ser menor porque corresponde a cada parámetro muestral.

De igual manera para Y_2, X_2

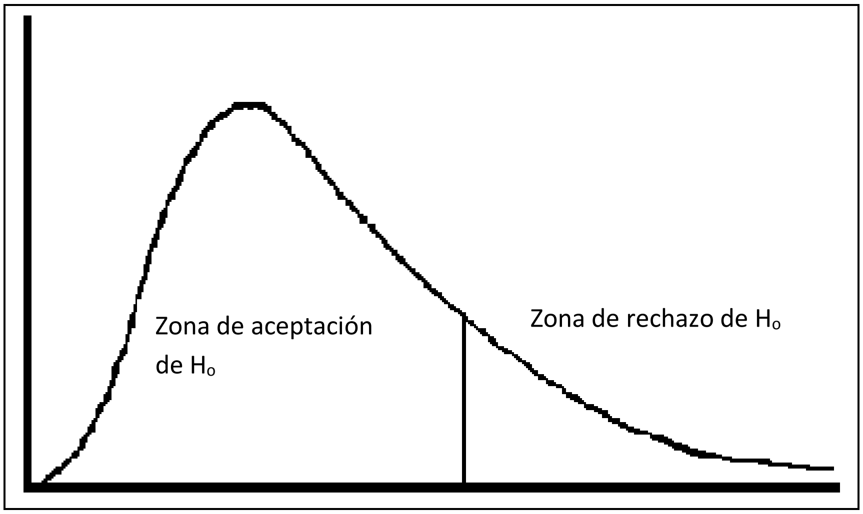
Dependent Variable: Y2				
Method: Least Squares				
Sample(adjusted): 1 10				
Included observations: 10 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.278963	3.443383	0.952250	0.3689
X2	0.834637	0.096213	8.674885	0.0000
R-squared	0.903908	Mean dependent var		32.75000
Adjusted R-squared	0.891897	S.D. dependent var		5.401080
S.E. of regression	**1.775825**	Akaike info criterion		4.163264
Sum squared resid	25.22845	Schwarz criterion		4.223781
Log likelihood	-18.81632	F-statistic		75.25363
Durbin-Watson stat	2.051248	Prob(F-statistic)		0.000024

Con los dos S.E. of regresión, elevándolos al cuadrado, calculamos las dos varianzas y F:

Se calcula $F = \dfrac{Varianza\ de\ residuos\ grandes}{Varianza\ de\ residuos\ pequeños} = \dfrac{3.154}{0.475} = 6.64$

Para calcular los grados de libertad, gl, de la F téorica Salvatore(1993:152) y Gujarati (1991:266) señalan que los grados de libertad tanto para el numerador como para el denominador se calculan con la fórmula: n-d- 2k/2, donde n= número de observaciones= 20, d= número de observaciones omitidas, en este caso ninguna, luego d=0, k= número de parámetros= 2 en cada grupo, luego tanto para el numerador como para el denominador gl= 20-0 -2(2)/2= 20-4/2= 8 gl

Así, la F teórica se obtiene en tablas para $\alpha = 1\%$ con 8 y 8 grados de libertad, y es $F_{\alpha} = 6.03 < F = 6.64,$ por lo que se rechaza la hipótesis de homocedasticidad y *se acepta que hay un problema de heterocedasticidad.* Gráficamente:



F_{α}=6.03

Con Eviews nuevamente los pasos serían los siguientes:

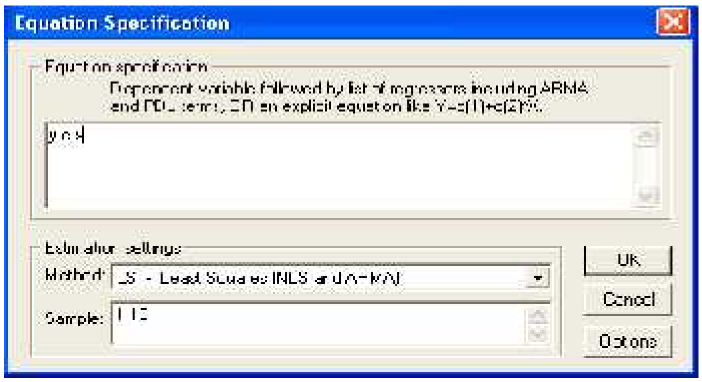
1. Clasificar en orden ascendente los valores de Y e X así: en la ventana del Workfile/Procs/Sort Series/ escribimos X, “ascending”/ok. Igual hacemos para Y y se obtienen los siguientes valores de X e Y:

obs	X	Y
1	6.2	6.1
2	8.1	8
3	10.3	10.3
4	12.1	12.1
5	14.1	13.1
6	16.4	14.8
7	18.2	17.9
8	20.1	19.8
9	22.3	19.9
10	24.1	21.6
11	28.3	25
12	26.1	25.5
13	30.1	29.3
14	32.3	31.2
15	36.6	31.8
16	34.5	33.1
17	38	33.5
18	44.7	38.6
19	40.2	38.8
20	42.3	40.7

2. Se elimina un determinado número de observaciones centrales para acentuar la variación entre los valores pequeños y grandes, en este caso, no se elimina ninguna observación.

3. Se corren las dos regresiones modificando la muestra (sample), así en Eviews: Quick/Estimate Ecuation/, para la muestra 1 el sample va de 1 a 10 y para la muestra 2 de 11 a 20:

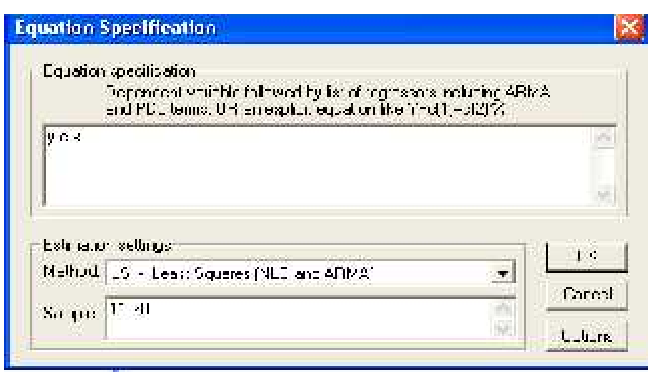
Para la primera muestra:



Cuya regresión es la siguiente:

Dependent Variable: Y				
Method: Least Squares				
Sample: 1 10				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.053316	0.616164	1.709474	0.1257
X	0.876016	0.037939	23.09013	0.0000
R-squared	0.985217		Mean dependent var	14.36000
Adjusted R-squared	0.983369		S.D. dependent var	5.346692
S.E. of regression	**0.689519**		Akaike info criterion	2.271210
Sum squared resid	3.803487		Schwarz criterion	2.331727
Log likelihood	-9.356051		F-statistic	533.1539
Durbin-Watson stat	1.745354		Prob(F-statistic)	0.000000

Para la segunda muestra:



Su regresión es:

Dependent Variable: Y				
Method: Least Squares				
Sample: 11 20				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.278963	3.443383	0.952250	0.3689
X	0.834637	0.096213	8.674885	0.0000
R-squared	0.903908	Mean dependent var		32.75000
Adjusted R-squared	0.891897	S.D. dependent var		5.401080
S.E. of regression	**1.775825**	Akaike info criterion		4.163264
Sum squared resid	25.22845	Schwarz criterion		4.223781
Log likelihood	-18.81632	F-statistic		75.25363
Durbin-Watson stat	1.831152	Prob(F-statistic)		0.000024

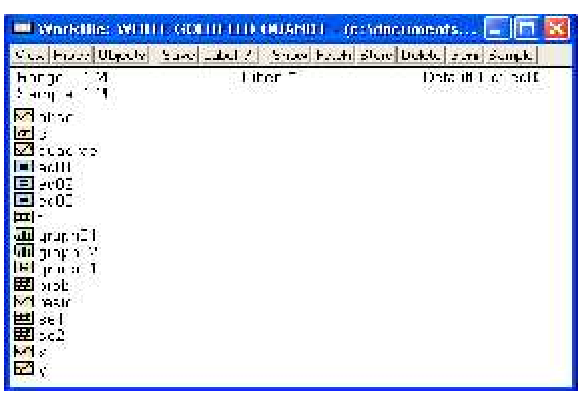
4. Ahora, almacenamos los S.E.of regresión así: para el primero en la ventana de comandos escribimos: scalar se1=@se/enter, mismo que aparecerá en el Workfile como se1, cuyo valor se ve dando doble click a dicha variable, el cual es mostrado como 0.689519 en la esquina inferior izquierda. Igualmente para almacenar el de la segunda muestra hacemos: scalar se2=@se/enter, cuyo valor es 1.775825

5. Generamos F escribiendo en la ventana de comandos: scalar f=(se2/se1)^2/enter y aparece el valor de F en la parte inferior izquierda de la pantalla haciendo doble click sobre el objeto escalar que acabamos de crear, el cual es F=6.632978, que es el mismo valor calculado manualmente.

6. Calculamos F_{α} con 8 y 8 grados de libertad tanto para el numerador como para el denominador y con α=5%. Para ello escribimos en la ventana de comandos: scalar prob=(1-@cfdist(f,8,8))/enter, cuyo valor es Probabilidad de F (Prob(F-statistic))=0.0074095

7. Dado que la probabilidad de F es menor que α=5%, se rechaza la hipótesis nula de homoscedasticidad y decimos que hay heteroscedasticidad entre las U_i de las veinte familias.

Comentarios: El Workfile definitivo que permitió obtener estos resultados se muestra a continuación:



VI.1.4.- Solución al problema de heteroscedasticidad.

¿Cómo se corrige o resuelve la heteroscedastidad?

Con opciones como las siguientes:

- La aplicación de la técnica de mínimos cuadrados generalizados y ponderados;
- La deflactación de los datos mediante algún índice apropiado;
- En general se recomienda transformar los valores en logaritmos, eliminar las variables responsables de la heteroscedasticidad o introducir variables binarias; Así veamos
- *La transformación de los datos en la forma funcional denominada logarítmica*.

VI.1.4.1.- Transformación de los datos en logaritmos, usando una forma funcional doble logarítmica.

Carrascal et al (2001) indica que las variables originales se transforman en logaritmos usando la forma funcional antes vista: log log, desarrollada por BOX-COX. La idea es que el logaritmo al tomar valores en un rango menor (ejemplo: si dos datos de la variable (X) son 100 y 10 pero sus logaritmos en base 10 son respectivamente 2 y 1), suaviza la dispersión que podrían mostrar los valores originales, lo cual es bueno; sin embargo **advierte** algo muy grave: que esta solución se basa en el supuesto de que el *modelo inicial es no lineal.* Si ello es cierto, se produce un modelo lineal en logaritmos, pero, si no lo es, comenta: " estimarlo en logaritmos no supone estimar el mismo modelo". Derivado de lo anterior es que se aconseja también intentar los otros métodos de solución como los arruna descritos..

Así pues, trabajando con esta opción se corre la regresión en forma doble logarítmica lineal. Usando los 20 datos originales y convirtiéndolos en logaritmos: usando Eviews vamos a *Quick/ Generate Series/enter equation,* ponemos lx=log(x); también ly=log(Y) y aparece la siguiente tabla:

obs	LX	LY
1	3.10458667847	2.99071973173
2	3.47506723023	3.44041809482
3	3.60004824041	3.45946628979
4	2.4932054526	2.4932054526
5	3.74478708605	3.70622809245
6	1.82454929205	1.80828877118
7	3.79997350162	3.65325227647
8	3.26193531433	3.23867845216
9	2.33214389524	2.33214389524
10	3.69386699562	3.65842024663
11	2.09186406168	2.07944154168
12	3.54095932404	3.49953328238

13	3.63758615973	3.51154543883
14	2.64617479738	2.57261223021
15	2.79728133483	2.69462718077
16	3.1822118405	3.07269331469
17	3.40452517175	3.37758751602
18	3.34286180465	3.21887582487
19	2.90142159408	2.88480071285
20	3.00071981507	2.9856819377

Vamos a *Quick:/* Estímate Equation: LY C LX/ ok

Se corre la regresión y se obtiene:

Log y = 0.0757 + 0.9562 log x R^2 =0.9935

(0.0574) (0.0183) RSS = 0.03757

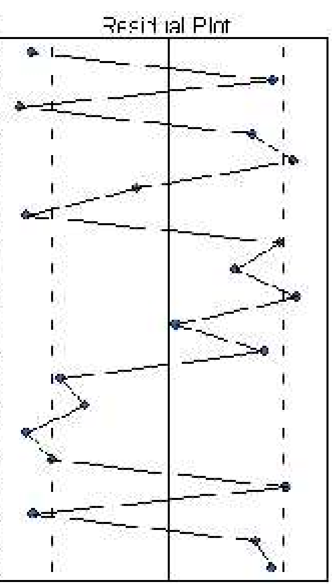
Dependent Variable: LY
Method: Least Squares
Sample: 1 20
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.075672	0.057393	1.318496	0.2039
LX	0.956186	0.018255	52.38022	0.0000

R-squared	0.993482	Mean dependent var	3.033911
Adjusted R-squared	0.993120	S.D. dependent var	0.550836
S.E. of regression	**0.045689**	Akaike info criterion	-3.239278
Sum squared resid	0.037575	Schwarz criterion	-3.139705
Log likelihood	34.39278	F-statistic	2743.688
Durbin-Watson stat	2.166013	Prob(F-statistic)	0.000000

Para calcular los residuos con E Views se estima la regresión, en el menú de View, seleccionar Actual, Fitted, Residual, después nos vamos a Actual Fitted, Table:

obs	Actual	Fitted	Residual
1	2.99072	3.04424	0.05352
2	3.44042	3.39846	0.04193
3	3.46947	3.51799	0.06852
4	2.49321	2.45964	0.03356
5	3.70623	3.65629	0.04994
6	1.80829	1.82028	-0.01199
7	3.65326	3.71916	-0.06590
8	3.22060	3.19409	0.04099
9	2.33214	2.30564	0.02651
10	3.08042	3.03770	0.05072
11	2.07944	2.07588	0.00356
12	3.46953	3.43149	0.03804
13	3.51155	3.55388	-0.04234
14	2.57261	2.60591	-0.03330
15	2.69463	2.75039	-0.05577
16	3.07269	3.11846	-0.04577
17	3.37759	3.33103	0.04655
18	3.21888	3.28207	-0.06320
19	2.88480	2.84997	0.03433
20	2.98568	2.97492	0.01076



Enseguida clasificamos las Ui en función de X, en los dos siguientes grupos:

Número de observación	Log Y_1 calculada	Log de x_1	Residuo u_i	Número de observación	Log Y_2 calculada	Log de x_2	Residuo u_i
6	1.8	1.82	-0.12	8	3.24	3.26	0.44
11	2.08	2.09	0.04	18	3.22	3.34	-0.53
9	2.33	2.33	0.27	17	3.38	3.40	0.47
4	2.49	2.49	0.34	2	3.44	3.48	0.42
14	2.57	2.65	-0.33	12	3.5	3.54	0.38
15	2.69	2.80	-0.56	3	3.46	3.60	-0.59
19	2.88	2.90	0.35	13	3.51	3.64	-0.42
20	2.99	3.00	0.41	10	3.66	3.69	0.51
1	2.99	3.20	-0.54	5	3.7	3.74	0.50
16	3.07	3.18	-0.46	7	3.65	3.80	-0.56

Una vez calculados los residuos de los dos grupos se corren sus dos regresiones y se obtiene, para el primero: Quick, Estimate Equation: LY- C- LX_1, ok

Dependent Variable: LY1
Method: Least Squares
Sample(adjusted): 1 10
Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.122770	0.083001	1.479135	0.1774
LX1	0.935596	0.031083	30.09966	0.0000

R-squared	0.991247	Mean dependent var	2.589000
Adjusted R-squared	0.990153	S.D. dependent var	0.422518
S.E. of regression	**0.041927**	Akaike info criterion	-3.328910
Sum squared resid	0.014063	Schwarz criterion	-3.268393
Log likelihood	18.64455	F-statistic	905.9898
Durbin-Watson stat	1.786700	Prob(F-statistic)	0.000000

Y para el segundo grupo: Quick Estimate Equation: LY_2 –C- LX_2, , ok

Dependent Variable: LY2
Method: Least Squares
Sample(adjusted): 1 10
Included observations: 10 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.320335	0.358071	0.894614	0.3971
LX2	0.889170	0.100780	8.822901	0.0000

R-squared	0.906807	Mean dependent var	3.476000
Adjusted R-squared	0.895158	S.D. dependent var	0.166012
S.E. of regression	0.053754	Akaike info criterion	-2.831958
Sum squared resid	0.023116	Schwarz criterion	-2.771441
Log likelihood	16.15979	F-statistic	77.84358
Durbin-Watson stat	2.189455	Prob(F-statistic)	0.000021

Se dice que hay una solución porque se observa que no hay un aumento significativo en el valor de los residuos (u_i) a medida que crecen los valores de x, es decir, se reduce la heterocedasticidad en las varianzas del error.

VI.1.4.1.1 Aplicación de F.

De las dos regresiones anteriores tenemos: con los cálculos de Maddala:

Grupo 1	Grupo 2
log y = 0.122 + 0.936x R^2 = 0.991; (0.083) (0.031) ±**=0.041927** $\hat{\pm}^2 = 0.001596$	log y = 0.320 + 0.889x R^2 = 0.907 (0.358) (0.100) ±**=0.053754** $\hat{\pm}^2 = 0.002789$

Así, $F = \dfrac{0.002789}{0.001596} = 1.75$; Como $F_{\sim} = 3.44$ para $\sim = 5\%$ y con α= 1% tenemos F teórica= 6.03 con 8 y 8 grados de libertad. En los dos casos vemos que no se rechaza la hipótesis de homoscedasticidad; se dice que desapareció la heteroscedasticidad, que los estimadores ahora son insesgados y eficientes y ratifican los motivos por los cuales en el capítulo IX se prefirió esta forma funcional.

En resumen, se debe señalar que es conveniente detectar si existe o no heteroscedasticidad, ya que de no identificarse este problema, ello ocasiona que:

a) Los estimadores de mínimos cuadrados sean ineficientes, aun cuando siguen siendo insesgados; es decir, cuando son ineficientes tienen una varianza más grande.
b) Los estimadores de las varianzas son sesgados. Ello nulifica (mejor dicho, altera los resultados de) las pruebas de significación que se realizan para probar la bondad de los estimadores.
c) Se relaja el supuesto de que la varianza del término de error (u_i) es constante.

VI.1.5.- Ejemplos utilizando EViews 5.

La ilustración al tema de heteroscedasticidad se realizará mediante EViews 5 utilizando datos actuales de la economía mexicana.

Ejemplo 1. Consumo Privado en función del Producto Interno Bruto.

Utilizaremos el Producto Interno Bruto y el Consumo Privado a precios constantes de 2003. Las serien poseen 124 observaciones, las cuales comprenden desde el primer trimestre de 1980 hasta el cuarto trimestre de 2010. Para comenzar el análisis de heteroscedasticidad como primer paso se generará la ecuación de regresión como se puede apreciar en el Cuadro 1. La regresión se realiza bajo el supuesto que no existe heteroscedasticidad, para hacer más sencilla la

explicación haremos uso de la función donde el Consumo Privado depende del Producto Interno Bruto.
El modelo planteado es el siguiente:

$$CP = f\ (PIB)$$

Para realizar la regresión en EViews en la barra de comandos escribimos: ls cp c PIB, seguido de oprimir la tecla enter, con lo cual se desplegará una ventana como la que se muestra en el Cuadro 1.

Cuadro 1. Resultados de la ecuación de regresión.

Dependent Variable: CP
Method: Least Squares
Date: 09/15/11 Time: 17:31
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.92E+08	75189357	-5.216670	0.0000
PIB	0.720557	0.011450	62.93031	0.0000

R-squared	0.970114	Mean dependent var	4.22E+09
Adjusted R-squared	0.969869	S.D. dependent var	1.09E+09
S.E. of regression	1.89E+08	Akaike info criterion	40.96583
Sum squared resid	4.35E+18	Schwarz criterion	41.01132
Log likelihood	-2537.881	F-statistic	3960.224
Durbin-Watson stat	0.653227	Prob(F-statistic)	0.000000

De acuerdo a los resultados de la estimación, la ecuación de regresión es:

$$\hat{CP} = -392238091.6 + 0.7205 * PIB$$

La interpretación de los coeficientes es la siguiente si el PIB crece en mil pesos constantes de 2003, el Consumo Privado se incrementará en 0.7205 miles de pesos constantes de 2003, y viceversa, si el PIB disminuye en mil pesos constantes de 2003, el consumo privado disminuirá en 0.7205 miles de pesos constantes de 2003.

En cuanto al valor que asume el R^2 es de 0.970114, lo que significa que el 97.0114% de las variaciones en el Consumo Privado se explican por variaciones en el PIB.

Dentro de la pantalla de resultados de la ecuación (ver Cuadro 1) existe un método de detección no formal para determinar la existencia o no de heteroscedasticidad en el modelo. Este método consiste en que el error estándar de la regresión (S.E. of regression) debe ser bajo cuando la varianza de los errores es mínima y constante. En este caso el valor que arroja es de 188, 758, 137.1797 por lo que en un primer acercamiento podríamos concluir que probablemente exista heteroscedasticidad, debido a que el error estándar de la regresión es muy grande.

La detección de heteroscedasticidad utilizando EViews se puede realizar por dos métodos: 1. Métodos informales o gráficos; y 2. Métodos formales o numéricos. En cuanto al primero tenemos: La gráfica de residuales, la representación gráfica del valor absoluto de los errores y la representación gráfica del cuadrado de los errores. Esto nos indica el grado de la heteroscedasticidad al observar la dispersión, pudiendo ser ésta constante o variable. En caso de ser variable puede ser reflejo de las dispersiones en aumento para la perturbación aleatoria y por tanto indicio de la existencia de heteroscedasticidad.

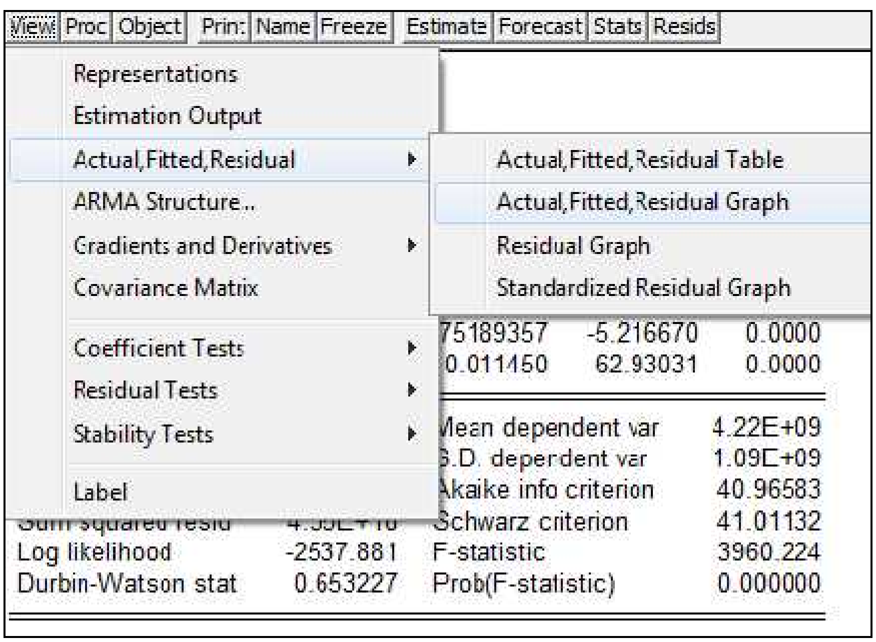
Método gráfico.

Gráfica de residuos.

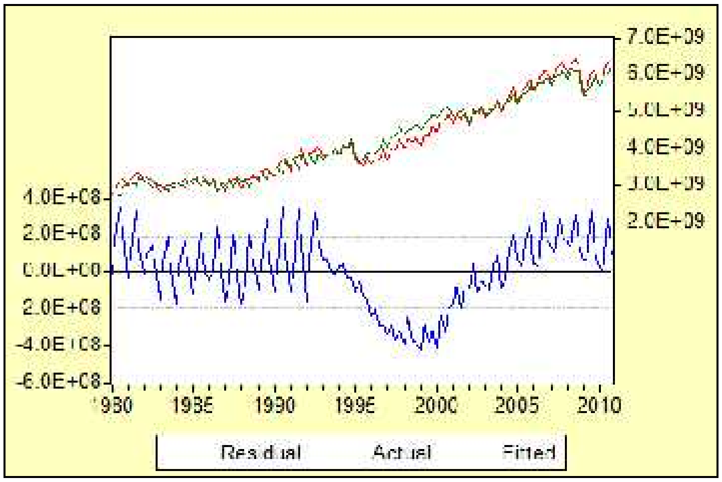
Se obtendrá la gráfica de los valores observados y estimados de la variable dependiente y sus residuos para determinar si estos últimos presentan heteroscedasticidad.

Así, en la ventana de la ecuación de regresión vamos a: View/ Actual, Fitted, Residual/ Actual, Fitted, Residual Graph (ver Cuadro 2), con lo cual se obtendrá una ventana como la contenida en el Cuadro 3.

Cuadro 2. Generación de la gráfica de valores observados y estimados de la variable dependiente y los errores.



Cuadro 3. Representación de los valores observados y estimados de la variable dependiente y los errores.



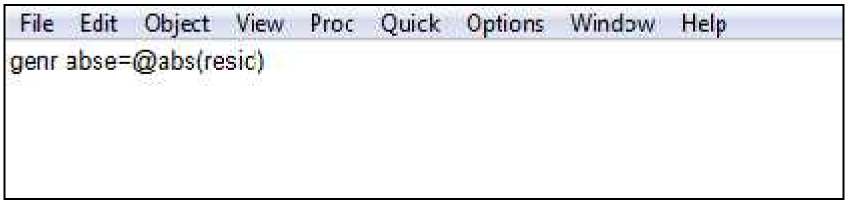
En cuanto a los errores (Residual), observamos que éstos se salen de las bandas de confianza, ello es muestra de que los errores no poseen la misma varianza, por lo que se concluye, mediante el análisis gráfico, que el modelo presenta heteroscedasticidad. Asimismo, nótese que claramente se pueden visualizar tres periodos en cuanto al comportamiento de los residuos: 1) entre 1980 y 1993; 2) entre 1993 y 2003; y 3) entre 2003 y 2010. Ello denota que los errores no tienen un mismo comportamiento, por lo que no poseen la misma o similar varianza, lo cual demuestra, visualmente, que el modelo presenta heteroscedasticidad.

1. **Diagrama de dispersión entre los valores absolutos de los errores, el cuadrado de los mismos y la variable independiente.**

A continuación se procederá a realizar el diagrama de dispersión entre los errores en valores absolutos y la variable independiente, en este caso el PIB, para determinar la presencia de heteroscedasticidad o no en el modelo. Dicho procedimiento se lleva a cuando se trata de regresión simple, como en este caso. Cuando se trate de regresión múltiple existen dos opciones: 1) Realizar los diagramas de dispersión entre cada una de las variables independientes incluidas en el modelo frente a los errores en términos absolutos y al cuadrado; y 2) Realizar los diagramas de dispersión entre los valores estimados de la variable dependiente frente a los errores en términos absolutos y al cuadrado.

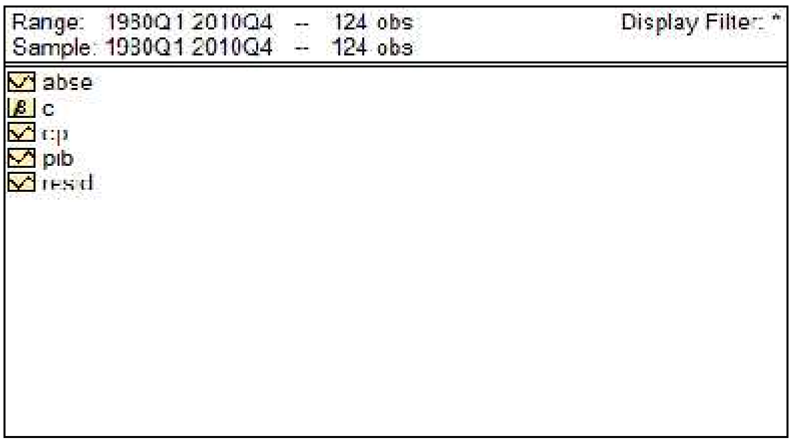
Así, para generar el valor absoluto de los términos de error en EViews nos situamos dentro de la barra de comandos, en ella escribiremos el comando “genr abse=@abs(resid)” y enseguida enter. Esto se muestra a continuación en el Cuadro 4.

Cuadro 4. Barra de Comandos.



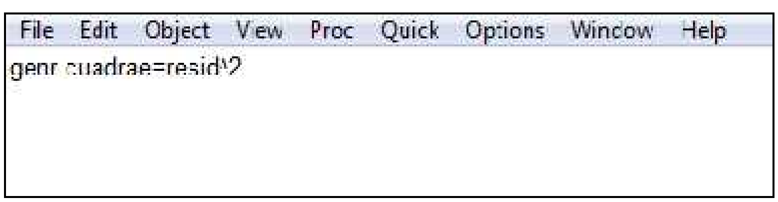
En la pantalla del Workfile, como se aprecia en el Cuadro 5, aparece la nueva serie generada con el procedimiento anterior, el cual ostenta el nombre de: “abse”.

Cuadro 5. Archivo de Trabajo (Workfile).



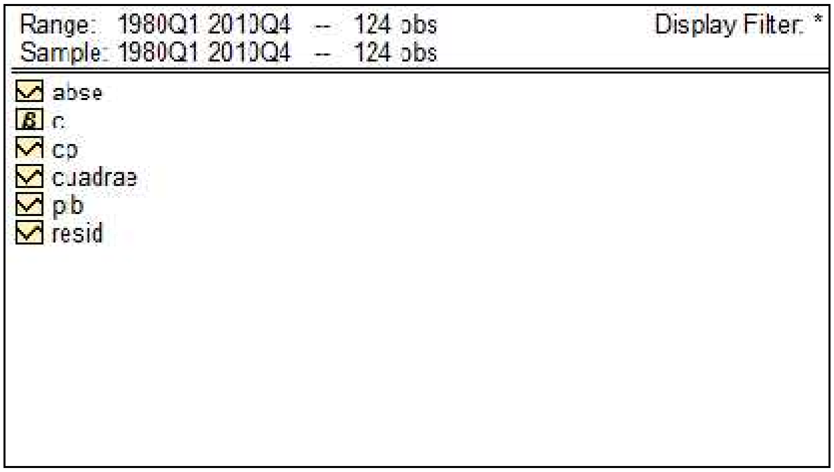
Para generar el cuadrado de los errores, nuevamente nos situamos dentro de la barra de comandos, en ella escribiremos el comando siguiente: “genr cuadrae=resid^2” y enseguida enter. Esto se muestra a continuación en el Cuadro 6.

Cuadro 6. Barra de Comandos.



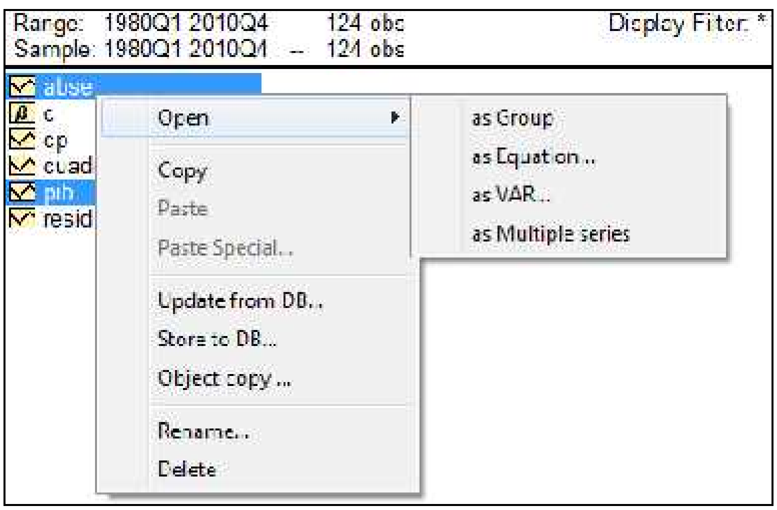
En la pantalla de Workfile como se aprecia en el Cuadro 7, aparece el comando anteriormente generado, mismo que presenta el nombre de “cuadrae”.

Cuadro 7. Archivo de Trabajo (Workfile).



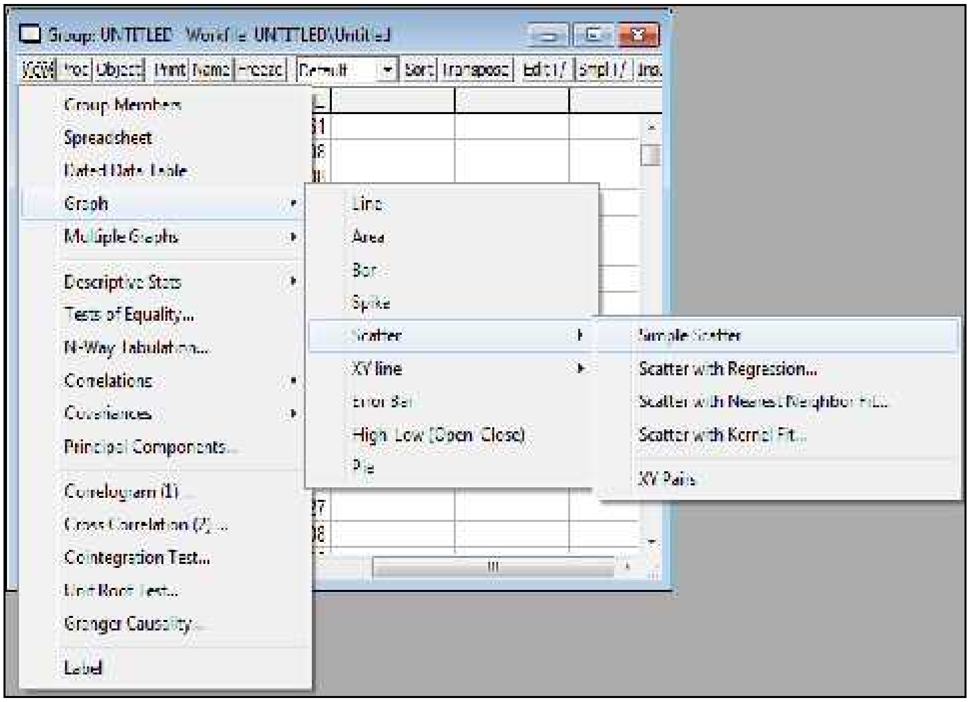
Una vez generadas ambas series se seleccionan el "pib" y "abse" se presiona CTRL y se da clic izquierdo, aparecerá una pantalla, se selecciona "open", en seguida sale otra ventana donde se seleccionará "as a grup", donde aparecerá una nueva ventana que solo incluye las variables previamente seleccionadas. Para fines ilustrativos esto se puede observar en el Cuadro 8.

Cuadro 8. Pasos para abrir un grupo.



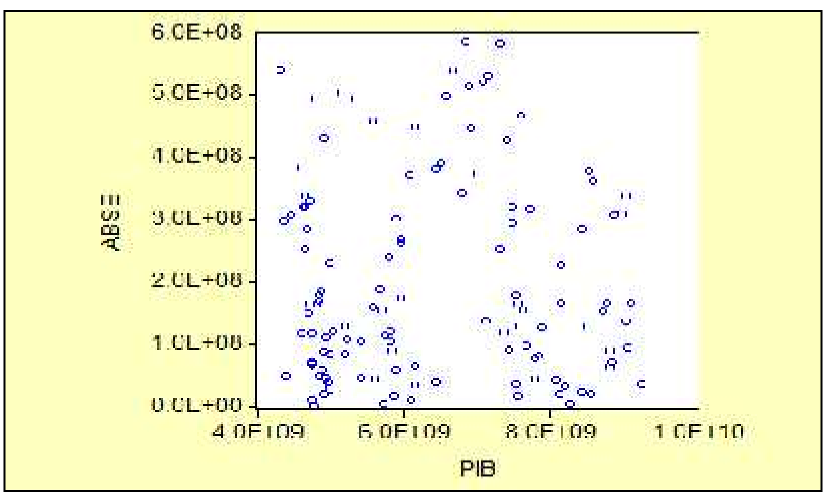
En la nueva ventana que se genera vamos a: View/Graph/Scatter/Simple Scatter de la barra de herramientas del grupo, tal y como se visualiza en el Cuadro 9.

Cuadro 9. Dispersión sencilla.



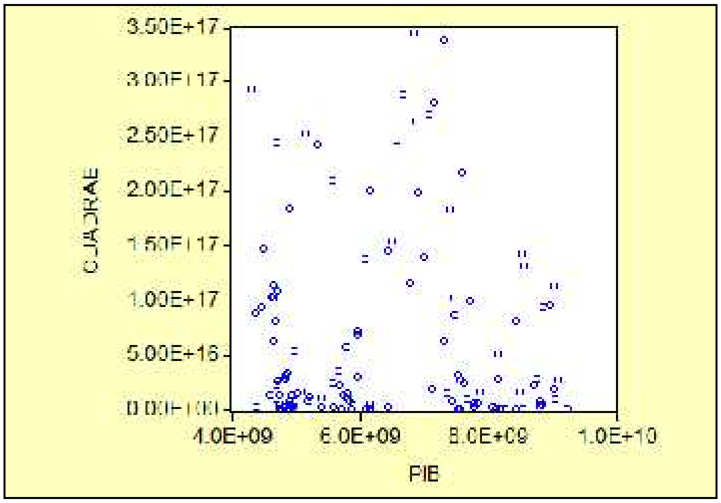
En seguida aparecerá la gráfica del valor absoluto de los errores de la ecuación frente al PIB (véase Cuadro 10). En el gráfico del valor absoluto de los errores de la ecuación frente a la variable independiente se aprecia que a medida que aumentan los valores del PIB la dispersión de los errores aumenta haciendo determinante la presencia de heteroscedasticidad.

Cuadro 10. Diagrama de dispersión del valor absoluto de los errores de la ecuación.



Ahora bien, para generar el gráfico del valor del cuadrado de los errores frente al PIB se sigue un procedimiento similar, sólo que al abrirse el grupo se seleccionará "cuadrae" en lugar de "abse". Al igual que el gráfico del valor absoluto de los errores, el gráfico del valor del cuadrado de los errores muestra que a medida que aumentan los valores del PIB la dispersión de los errores al cuadrado también aumenta, aunque no están tan dispersos los errores como en el gráfico anterior (ver Cuadro 10), por lo que concluimos que existe heteroscedasticidad en el modelo.

Cuadro 11. Diagrama de dispersión entre el valor del cuadrado de los errores y la variable independiente.



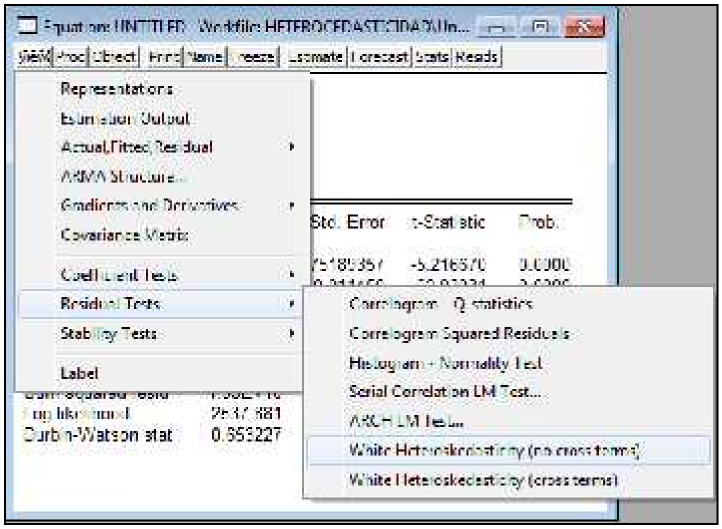
Método formal.

Por lo que respecta a las pruebas formales de heteroscedasticidad, en EViews deberemos abrir la ventana de la regresión original que se realizó desde un inicio, es decir, el Consumo Privado en función del Producto Interno Bruto (PIB). En este caso se utilizará la prueba de White tanto para datos no cruzados como para datos cruzados.

Para realizar la prueba White para datos no cruzados[20] en la barra de herramientas de la pantalla de regresión (ver Cuadro 1) iremos a: View/Residual Test/White Heteroskedasticity (no cross terms) (Véase Cuadro 12).

[20]Se refiere a términos no cruzados cuando se realiza la regresión de los errores obtenidos en la regresión original al cuadrado tomando como explicativas todas las variables independientes de la regresión inicial y sus valores al cuadrado.

Cuadro 12. Ventana para generar la prueba de Heteroscedasticidad de White para datos no cruzados.



En seguida aparecerá una ventana que muestra los valores de la prueba de White para datos no cruzados, a continuación se muestran los más importantes:

- El estadístico F, que es una prueba para variables omitidas que evalúa la significación conjunta de todos los productos cruzados excluyendo la constante.

- El estadístico $\lambda = n * R^2$, donde n es el tamaño de la muestra y R^2 es el coeficiente de determinación de la regresión auxiliar. Este estadístico se distribuye asintóticamente, bajo la hipótesis nula, como una χ^2.

Las variables incluidas en la regresión auxiliar no deberían tener ningún poder explicativo sobre los residuos al cuadrado, por lo cual, el R^2 debe ser pequeño. Asimismo, si el valor muestral del estadístico es alto se rechazar la hipótesis nula, ya que el p-value es menor al 5% por lo cual el modelo presenta homoscedasticidad.

Se plantea la siguiente prueba de hipótesis tanto para la prueba White de datos cruzados y datos no cruzado:

$H_0 : \pm^2_{ui} = \pm^2_{uj}$; *las varianzas de los errores son iguales presentándose homoscedasticidad.*

$H_0 : \pm^2_{ui} \neq \pm^2_{uj}$; *las varianzas de los errores son diferentes presentándose no homoscedasticidad.*

El valor obtenido o calculado de “F” se contrasta con la “F_α”. Para encontrar el valor de la “F_α” se utiliza un nivel de significación de 5%, y los grados de libertad para el numerador y el denominador. Los grados de libertad en el numerador se determinan con k-1=3-1=2, y los grados de libertad en el denominador con n-k=124-3=121.

Así, la estadística "F_{α}" teórica es:

$$F_{\alpha} = 3.0711$$

La estadística "F" calculada se contrasta con el valor de la estadística "F_{α}" para ello se establecen los siguientes criterios de decisión:

Si "F" calculada es menor o igual que la "F" teórica no se rechaza la Ho.

Si "F" calculada es mayor a "F" teórica se rechaza la Ho.

Como "F" calculada es menor que la "F" teórica, 2.6686 < 3.0711, no se rechaza la Ho, y por tanto, concluimos que con esta prueba el modelo presenta homoscedasticidad.

Este resultado puede ser corroborado mediante la probabilidad asociada a la estadística "F" que el programa proporciona como "Prob (F-statistic)", ésta indica la probabilidad mínima a la cual se rechazaría la Ho. Así, los criterios de decisión son:

Si la probabilidad asociada a la estadística "F(White)" es menor o igual a 0.05 se rechaza la Ho.

Si la probabilidad asociada a la estadística "F(White)" es mayor a 0.05 no se rechaza la Ho.

Por lo que, al ser la probabilidad asociada a la estadística "F (White)" mayor a 0.05 (ver Cuadro 13) no se rechaza la Ho y se corrobora la conclusión anterior en la que el modelo presenta homoscedasticidad. Este resultado se obtiene debido a que la constante y el cuadrado del PIB no explican de forma individual a los residuos al cuadrado y de forma conjunta tampoco, ya que no son estadísticamente significativos.

En cuanto a la regresión auxiliar tenemos que: R^2 = 0.042246 = 4.2246%, por lo que presenta un valor pequeño, lo cual significa que las variables incluidas en esta regresión no tienen poder explicativo sobre los residuos al cuadrado. Con ello se corrobora la ausencia de heteroscedasticidad.

Cuadro 13. Prueba de Heteroscedasticidad de White para datos no cruzados.

White Heteroskedasticity Test:

F-statistic	2.668641	Probability	0.073427
Obs*R-squared	5.230540	Probability	0.072056

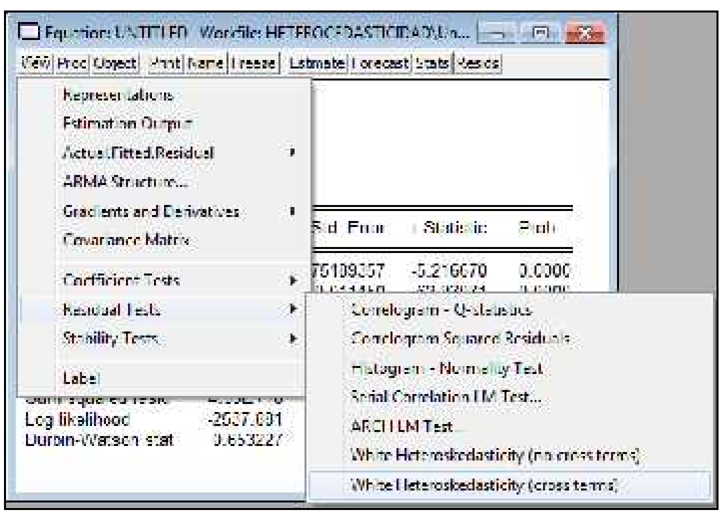
Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 09/16/11 Time: 17:42
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.69E+17	9.71E+16	1.736142	0.0851
PIB	61017500	30429047	2.005106	0.0472
PIB^2	-0.004331	0.002283	-1.897348	0.0602

R-squared	0.042246	Mean dependent var	3.51E+16
Adjusted R-squared	0.026416	S.D. dependent var	4.44E+16
S.E. of regression	4.38E+16	Akaike info criterion	79.49857
Sum squared resid	2.32E+35	Schwarz criterion	79.56680
Log likelihood	-4925.911	F-statistic	2.668641
Durbin-Watson stat	1.358620	Prob(F-statistic)	0.073427

Para realizar la prueba de White para datos cruzados, en la barra de herramientas de la pantalla de regresión, vamos a: View/Residual Test/White Heteroskedasticity (cross terms) (Véase Cuadro 14).

Cuadro 14. Ventana para generar la prueba de Heteroscedasticidad de White para datos cruzados.



Después de seguir los pasos para generar la prueba de heteroscedasticidad de White para datos cruzados, al igual que la prueba anterior utilizaremos la Prueba F, el p-value “Prob(F-statistic)” y el R^2 con el fin de determinar si en este caso se presenta heteroscedasticidad. Sin embargo, cabe mencionar que en regresión simple, como en este caso, los resultados que se obtienen con White con datos cruzados y no cruzados son los mismos.

Los valores de esta prueba se encuentran en el Cuadro 15, a continuación haremos un análisis de éstos:

- La “F_c” es menor que la “F_α”, 2.6686 < 3.0711, no se rechaza la Ho, por tanto, con esta prueba se concluye que el modelo presenta homoscedasticidad.
- La probabilidad asociada a la estadística “F(White)” es mayor a 0.05, por lo que no se rechaza la Ho.
- En la regresión auxiliar tenemos que: R^2 = 0.042246 = 4.2246%, como representa un valor pequeño, ello significa que las variables incluidas en esta regresión no tienen poder explicativo sobre los residuos al cuadrado y por tanto existe homoscedasticidad.

Cuadro 15. Prueba de Heteroscedasticidad de White para datos cruzados.

White Heteroskedasticity Test:

F-statistic	2.668641	Probability	0.073427
Obs*R-squared	5.238540	Probability	0.072856

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 09/16/11 Time: 17:44
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.69E+17	9.71E+16	-1.736142	0.0851
PIB	61017500	30429847	2.005186	0.0472
PIB^2	-0.004331	0.002233	-1.897348	0.0602

R-squared	0.042246	Mean dependent var	3.51E+16
Adjusted R-squared	0.026416	S.D. dependent var	4.44E+16
S.E. of regression	4.38E+16	Akaike info criterion	79.49857
Sum squared resid	2.32E+35	Schwarz criterion	79.56680
Log likelihood	-4925.911	F-statistic	2.668641
Durbin-Watson stat	1.358620	Prob(F-statistic)	0.073427

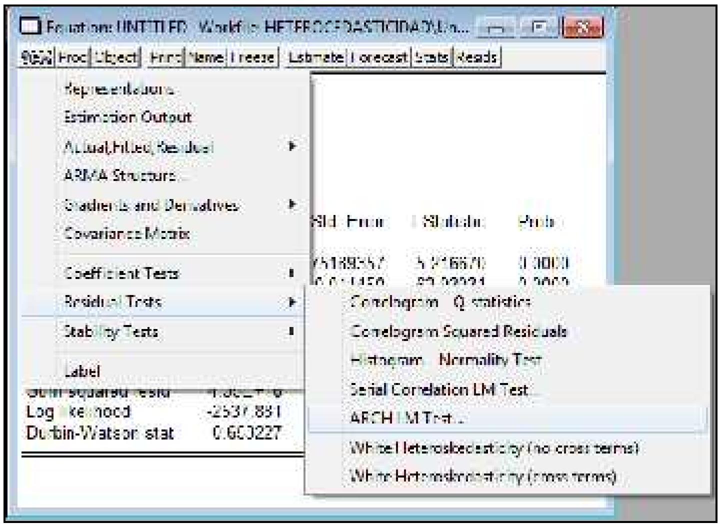
Es oportuno recalcar que los resultados para ambas pruebas, White con términos cruzados y no cruzados son iguales con y sin términos cruzados, esto no es un error al estimar los modelos, sino que se trata de un modelo de regresión simple, es decir, que el modelo sólo tiene un regresor. Para los modelos de regresión múltiple el resultado de ambas pruebas será diferente.

Otra prueba para determinar si el modelo presenta heteroscedasticidad es la prueba de Heteroscedasticidad Condicional Autorregresiva (ARCH).

La Prueba ARCH utiliza como variables explicativas de la regresión auxiliar los propios valores desplazados del residuo al cuadrado.

Para realizar esta prueba en EViews se abre nuevamente la pantalla de regresión original que hemos venido trabajando, en la barra de herramientas vamos a: View/Residual Test /ARCH LM Test/ número de rezagos de la varianza ARCH 1 (Véase Cuadro 16).

Cuadro 16. Ventana para la prueba ARCH LM.



Después aparecerá la ventana con la Prueba de ARCH (Véase Cuadro 15). En este caso se plantea la prueba de hipótesis siguiente:

H_0: Homoscedasticidad incondicional autorregresiva.
H_a: Heteroscedasticidad incondicional autorregresiva.

Para esta prueba también utilizaremos la prueba F y su p-value (Véase Cuadro 17). Recordando que el valor de "F_c" se contrasta con la "F_α". Para encontrar el valor de la "F_α" se utiliza el mismo nivel de significación, grados de libertad en el numerador y grados de libertad en el numerador:

$$\sim\ = 5\%$$

$$k - 1 = 2 - 1 = 1$$

$$n - k = 123 - 2 = 121$$

Así, la estadística "F_α" teórica es:

$$F_{1,102} = 3.9194$$

Para contrastar tanto "Fc" y "F_α" se mantienen los criterios de decisión:

Si "F" calculada es menor o igual que la "F" teórica no se rechaza la Ho.
Si "F" calculada es mayor a "F" teórica se rechaza la Ho.

Como “F” calculada es mayor que la “F” teórica, 20.77735 > 3.9194, se rechaza la Ho, y por tanto, concluimos que con esta prueba el modelo presenta no homoscedasticidad.

Mediante la probabilidad asociada a la estadística “F” que el programa proporciona como “Prob (F-statistic)” se corrobora la prueba de hipótesis no aceptada. También se siguen los criterios de decisión anteriores:

Si la probabilidad asociada a la estadística “F(ARCH)” es menor a 0.05 se rechaza la Ho.
Si la probabilidad asociada a la estadística “F” es mayor o igual a 0.05 no se rechaza la Ho.

La probabilidad asociada a la estadística “F(ARCH)” es menor a 0.05 (0.000022<0.05), se rechaza la Ho y se corrobora lo anterior, el modelo presenta no homoscedasticidad.

Cuadro 17. Prueba ARCH LM.

ARCH Test:

F-statistic	20.77735	Probability	0.000012
Obs*R-squared	18.02555	Probability	0.000022

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample (adjusted): 1980Q2 2010Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.19E+16	4.75E+15	4.597847	0.0000
RESID^2(-1)	0.382414	0.083896	4.558218	0.0000

R-squared	0.146549	Mean dependent var	3.53E+16
Adjusted R-squared	0.139496	S.D. dependent var	4.45E+16
S.E. of regression	4.12E+16	Akaike info criterion	79.37039
Sum squared resid	2.05E+35	Schwarz criterion	79.41612
Log likelihood	-4879.279	F-statistic	20.77735
Durbin-Watson stat	2.262002	Prob(F-statistic)	0.000012

Se podría concluir que los métodos gráficos utilizados para este modelo son un primer acercamiento del comportamiento de los términos de error y la distribución de su varianza. Pero puede que los gráficos no detecten el problema de forma tan clara, es por eso que se recurre a las pruebas sobre los residuales utilizando las más importantes.

En este caso a pesar de que la prueba de White para datos no cruzados y cruzados es la prueba más eficazpara la detección de heteroscedasticidad debido a que no se apoya en el supuesto de normalidad, es una distribución asintótica. Los resultados obtenidos con la prueba ARCH LM contradicen los resultados obtenidos con White. Por lo cual, a continuación se procederá a solucionar el problema.

Solución para heteroscedasticidad.

Como antes se indicó en el marco teórico del tema, existen diversas formas para solucionar la heteroscedasticidad como:

- La aplicación de la técnica de mínimos cuadrados ponderados: Los Mínimos cuadrados Ponderados (MCP) es una forma conveniente de obtener el estimador de mínimos cuadrados generalizados bajo heteroscedasticidad.

- La deflactación de los datos mediante algún índice apropiado: garantizará la homogenización de la serie y por tanto las varianzas serán constantes.

- La transformación de los datos en la forma funcional denominada logarítmica: con los logaritmos se logra una mayor suavización con respecto a los valores originales.

- Eliminar las variables responsables de la heteroscedasticidad.

- La introducción de variables dicotómicas (Dummy).

- Mínimos cuadrados generalizados: Este procedimiento consiste en transformar las variables originales de tal manera que satisfagan los supuestos del modelo clásico, para luego aplicar MCO a los mismos.

En este caso sólo se trabajará la corrección o eliminación de la heteroscedasticidad en el modelo propuesto en el primer ejemplo: el consumo privado está en función del Producto Interno Bruto, mediante la *transformación de los datos a logaritmos*, para verificarlo nuevamente con las pruebas de White y ARCH.

Para transformar las variables originales en logaritmos se inserta la siguiente indicación en la barra de comando: “genr LCP=LOG(CP)” y “genr LPIB=LOG(CPIB)”. Ambas variables aparecerán en la ventana de workfile con los nombres “LCP” y “LPIB”. Enseguida se corre la regresión de la forma ya habitual, obteniéndose los siguientes resultados:

Cuadro 18. Resultados de la nueva regresión.

Dependent Variable: LCP
Method: Least Squares
Date: 09/15/11 Time: 17:49
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.880167	0.396584	4.740905	0.0000
LPIB	1.064669	0.017584	60.54793	0.0000

R-squared	0.967793	Mean dependent var	22.13092
Adjusted R-squared	0.967529	S.D. dependent var	0.250100
S.E. of regression	0.045067	Akaike info criterion	3.345340
Sum squared resid	0.247785	Schwarz criterion	3.299852
Log likelihood	209.4111	F-statistic	3666.052
Durbin-Watson stat	0.841698	Prob(F-statistic)	0.000000

La ecuación de regresión es la siguiente:

$$L\hat{C}P = -1.880166522 + 1.064668988 * LPIB$$

Como puede verse en la ecuación anterior, la relación que se obtuvo entre el logaritmo del Producto Interno Bruto (LPIB) y del consumo privado (LCP) es directa o positiva. Asimismo, el valor de $\hat{a}$, que representa en este caso el consumo privado autónomo, es negativo, lo cual no concuerda con la teoría económica. Ello puede deberse a que probablemente el modelo obtenido no cumpla con todos los supuestos del método de estimación.

En cuanto a la interpretación de los coeficientes tenemos: si el Producto Interno Bruto se incrementa en un 1%, entonces el Consumo Privado aumenta en 1.0646%; y viceversa, si la Producto Interno Bruto disminuye en un 1%, el Consumo Privado disminuye en 1.0646%.

En cuanto a la bondad de ajuste, el R^2 es de 0.967793, es decir, el 96.7793% de los cambios en el Consumo Privado se explican por cambios en el Producto Interno Bruto, el resto es explicado por otras variables que no fueron incluidas en el modelo.

Para comprobar la usencia de heteroscedasticidad en el nuevo modelo planteado se procede a realizar las pruebas White y ARCH.
En primer lugar se genera la prueba de White para datos no cruzados, como ya se había explicado previamente, utilizando la regresión con logaritmos (véase Cuadro 18).

Se procede hacer el análisis de la prueba F, se debe encontrar el valor de la "F_{α}" utilizando:

$$\alpha = 5\%$$
$$k-1 = 3-1 = 2$$
$$n-k = 124-3 = 121$$

Así, la estadística “F_α” teórica es:

$$F_\alpha = 3.0711$$

La estadística “Fc” se contrasta con el valor de la “F_α” para ello se establecen los siguientes criterios de decisión:

Si “F” calculada es menor o igual que la “F” teórica no se rechaza la Ho.
Si “F” calculada es mayor a “F” teórica se rechaza la Ho.

Como “F” calculada es menor que la “F” teórica, 1.265781 < 3.0711, no se rechaza la Ho, concluimos que con esta prueba el modelo presenta homoscedasticidad.

Este resultado puede ser corroborado mediante la probabilidad asociada a la estadística “F” que el programa proporciona como “Prob (F-statistic)” indica la probabilidad mínima a la cual se rechazaría la Ho. Así, los criterios de decisión son:

Si la probabilidad asociada a la estadística “F” es menor a 0.05 se rechaza la Ho.
Si la probabilidad asociada a la estadística “F” es mayor o igual a 0.05 no se rechaza la Ho.

Por lo que, al ser la probabilidad asociada a la estadística “F” mayor a 0.05, no se rechaza la Ho debido a que 0.285726 > 0.05 y se corrobora la conclusión anterior en la que el modelo presenta homoscedasticidad.

Por último dentro de la regresión auxiliar tenemos que: $R^2 = 0.020493 = 2.0493\%$, presenta un valor pequeño lo que significa que las variables incluidas en esta regresión no tienen poder explicativo sobre los residuos al cuadrado y por tanto el nuevo modelo planteado presenta homoscedasticidad.

Para esta prueba fue favorable la transformación de las variables tanto independiente como dependiente a logaritmos. Esto debido a la comparación de la prueba White para datos no cruzados de la regresión original (ver Cuadro 1) y de la prueba White para datos no cruzados que se utilizó con la transformación a logaritmos (ver Cuadro 18), concluyendo que las cuatro pruebas concuerdan en que este modelo se comporta de manera homoscedástica.

Cuadro 19. Prueba de Heteroscedasticidad de White para datos no cruzados.

White Heteroskedasticity Test:

F-statistic	1.265781	Probability	0.285726
Obs*R-squared	2.541161	Probability	0.280669

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.006313	0.006339	-0.995981	0.3212
CP	4.45E-12	2.94E-12	1.515131	0.1323
CP^2	-5.03E-22	3.23E-22	-1.555867	0.1224

R-squared	0.020493	Mean dependent var	0.002939
Adjusted R-squared	0.004303	S.D. dependent var	0.003527
S.E. of regression	0.003520	Akaike info criterion	-8.437031
Sum squared resid	0.001499	Schwarz criterion	-8.368798
Log likelihood	526.0959	F-statistic	1.265781
Durbin-Watson stat	1.026502	Prob(F-statistic)	0.285726

Se siguen los mismos pasos para generar la prueba de heteroscedasticidad de White para datos cruzados sobre la regresión con logaritmos (Véase Cuadro 19).

En este caso se hará un análisis de los valores estadísticos de esta prueba:

- La "F_c" es menor que la "F_α", 0.9683 < 3.0711, no se rechaza la Ho, por tanto, con esta prueba se concluye que el modelo no presenta problemas de heteroscedasticidad.

- La probabilidad asociada a la estadística "F_α" es mayor a 0.05 no se rechaza la Ho.

- En la regresión auxiliar tenemos que: R^2 = 0.020493 = 2.0493%, presenta un valor pequeño lo que significa que las variables incluidas en esta regresión no tienen poder explicativo sobre los residuos al cuadrado y por tanto hay homoscedasticidad.

Cuadro 20. Prueba de Heteroscedasticidad de White para datos cruzados.

White Heteroskedasticity Test:

F statistic	0.968383	Probability	0.382620
Obs*R-squared	1.953516	Probability	0.376530

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.876146	2.774285	-0.676263	0.5002
LPIB	0.167631	0.245895	0.681717	0.4967
LPIB^2	-0.003740	0.005448	-0.686446	0.4937

R-squared	0.015754	Mean dependent var	0.001998
Adjusted R-squared	-0.000514	S.D. dependent var	0.002459
S.E. of regression	0.002459	Akaike info criterion	-9.154084
Sum squared resid	0.000732	Schwarz criterion	-9.085851
Log likelihood	570.5532	F-statistic	0.968383
Durbin-Watson stat	1.709392	Prob(F-statistic)	0.382620

Como se comentó anteriormente, los resultados para ambas pruebas de White, términos cruzados y no cruzados, son iguales con y sin términos cruzados, esto no es un error al estimar los modelos, sino que se trata de un modelo de regresión simple, es decir, que el modelo sólo tiene un regresor. Para los modelos de regresión múltiple el resultado de ambas pruebas será diferente.

Para la última prueba seguimos los pasos anteriormente descritos para poder generar la Prueba ARCH (véase Cuadro 21).

Cuadro 21. Prueba ARCH LM.

ARCH Test:

F-statistic	3.160484	Probability	0.077953
Obs*R-squared	3.130944	Probability	0.076819

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample (adjusted): 1980Q2 2010Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001691	0.000285	5.939604	0.0000
RESID^2(-1)	0.159501	0.089720	1.777775	0.0780

R-squared	0.025455	Mean dependent var	0.002012
Adjusted R-squared	0.017401	S.D. dependent var	0.002464
S.E. of regression	0.002443	Akaike info criterion	-9.175400
Sum squared resid	0.000722	Schwarz criterion	-9.129674
Log likelihood	566.2871	F-statistic	3.160484
Durbin-Watson stat	2.097915	Prob(F-statistic)	0.077953

Así como para la prueba de ARCH LM anterior, en esta prueba también utilizaremos la prueba F y su p-value. Recordando que el valor de “F_c” se contrasta con la “F_α”. Para encontrar el valor de la “F_α” se utiliza el mismo nivel de significación, grados de libertad en el numerador y grados de libertad en el numerador:

$$\alpha = 5\%$$
$$k-1=2-1=1$$
$$n-k=123-2=121$$

Así, la estadística “F_α” teórica es:

$$F_\alpha = 3.9194$$

Para contrastar tanto “Fc” y “F_α” se mantienen los criterios de decisión:

Si “F” calculada es menor o igual que la “F” teórica no se rechaza la Ho.

Si “F” calculada es mayor a “F” teórica se rechaza la Ho.

Como “F” calculada es menor que la “F” teórica, 3.160484 < 3.9194, no se rechaza la Ho, y por tanto, concluimos que con esta prueba el nuevo modelo presenta homoscedasticidad.

Mediante la probabilidad asociada a la estadística “F” que el programa proporciona como “Prob (F-statistic)” se corrobora la prueba de hipótesis. También se siguen los criterios de decisión anteriores:

Si la probabilidad asociada a la estadística “F” es menor a 0.05 se rechaza la Ho.

Si la probabilidad asociada a la estadística “F” es mayor o igual a 0.05 no se rechaza la Ho.

La probabilidad asociada a la estadística “F” es mayor a 0.05 (0.077953>0.05), por lo tanto no se rechaza la Ho y se corrobora la conclusión anterior, a saber, que el nuevo modelo presenta homoscedasticidad.

Así, concluimos que la transformación de las variables originales en logaritmos es un método útil y confiable para solucionar el problema de la heteroscedasticidad.

Ejemplo 2. Formación Bruta de Capital en función del Consumo Gubernamental, Consumo Privado e Importaciones.

Para llevar a cabo este modelo se utilizarán las variables siguientes: Formación Bruta de Capital (FBK) como variable dependiente y como variables regresoras Consumo Gubernamental (CG), Consumo Privado (CP) e Importaciones (M). La información se obtuvo de INEGI para el periodo 1980-2010 en lapsos trimestrales con el fin de explicar la actividad económica mexicana.

En este sentido el modelo planteado es el siguiente:

$$FBK = f(CG, CP, M)$$

La regresión se corre en EViews 5 siguiendo los pasos ya habituales y conocidos para realizarla, obteniéndose los siguientes resultados (ver Cuadro 22):

Cuadro 22. Resultados de la regresión.

Dependent Variable: FBK
Method: Least Squares
Regresión Original
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.11E+08	41072384	-12.44825	0.0000
CG	0.039960	0.063860	0.625744	0.5327
CP	0.392314	0.008569	45.78052	0.0000
M	0.859517	0.067529	12.72810	0.0000

R-squared	0.969350	Mean dependent var	1.21E+09
Adjusted R-squared	0.968584	S.D. dependent var	4.15E+08
S.E. of regression	73501593	Akaike info criterion	39.09524
Sum squared resid	6.48E+17	Schwarz criterion	39.18622
Log likelihood	-2419.505	F-statistic	1265.071
Durbin-Watson stat	0.746549	Prob(F-statistic)	0.000000

Es necesario plantear la prueba de hipótesis con la que se tomará la decisión de presencia de heteroscedasticidad u homoscedasticidad a lo largo de éste tema:

$$Ho: Homoscedasticidad.$$
$$Ha: Heteroscedasticidad.$$

Del Cuadro 22 y las hipótesis planteadas, se puede concluir que el modelo puede presentar heteroscedasticidad debido a que el error estándar de la regresión, presentado por S.E. of regression y Sum squared resid, que representa el ajuste de los errores a la constante, presentan valores muy altos, 73501693 y 6.48×10^{17} respectivamente, lo que puede indicar en primera instancia no homoscedasticidad, es decir, que hay heteroscedasticidad.

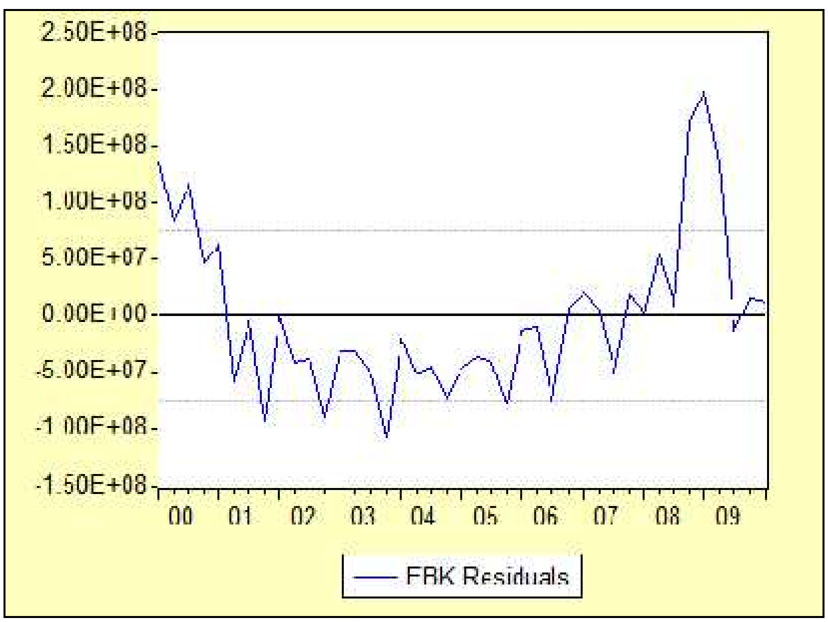
A continuación se aplicarán otros métodos para determinar la presencia o ausencia de la heteroscedasticidad en el modelo.

Método gráfico.

Gráfica de residuos.

En primera instancia se realizará el gráfico de los residuales (ver Cuadro 23) siguiendo los pasos descritos en el ejemplo 1 de este tema para obtenerlo con EViews.

Cuadro 23. Residuales.

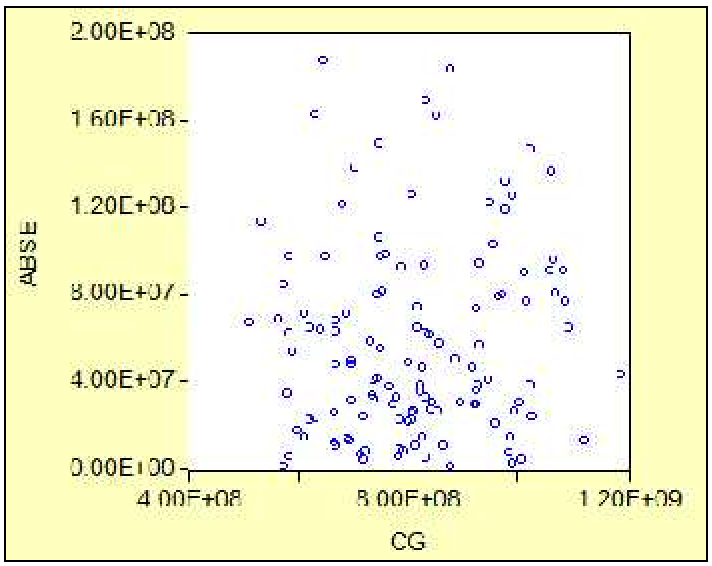


En éste Cuadro se observa que la serie de los residuales se sale de las bandas de confianza en los primeros años de la serie y en los últimos, principalmente, aunque también se sale en los años intermedios. Derivado de lo anterior se puede afirmar que la serie tiene varianzas diferentes, por lo que el modelo presenta heteroscedasticidad.

Diagrama de dispersión entre los valores absolutos de los errores, el cuadrado de los mismos y la variable independiente.

Ahora, para identificar la presencia o ausencia de heteroscedasticidad se utilizarán diagramas de dispersión de los valores absolutos y los valores al cuadrado de los errores frente a los valores que asumen cada una de las variables independientes. Siguiendo los pasos (comandos) utilizados en el ejemplo anterior para generar los valores absolutos y el cuadrado de los residuales, así como la generación de los diagramas correspondientes obtenemos lo siguiente:

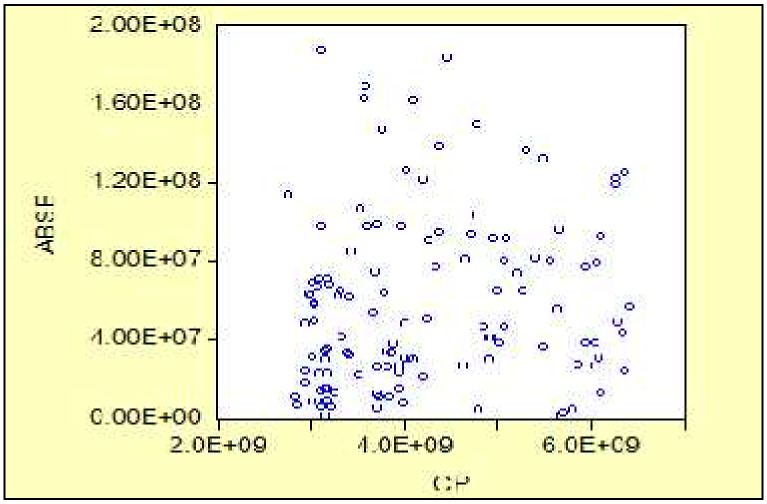
Cuadro 24. Diagrama de dispersión de valores absolutos de los errores frente al Consumo Gubernamental (CG).



Por lo que respecta al análisis del Cuadro 24 se puede observar que conforme aumenta el valor del CG los valores de los errores en términos absolutos se dispersan más, por lo que se puede concluir que se está ante la presencia de heteroscedasticidad.

Utilizando el mismo método se obtiene el Cuadro siguiente que analiza a la variable CP:

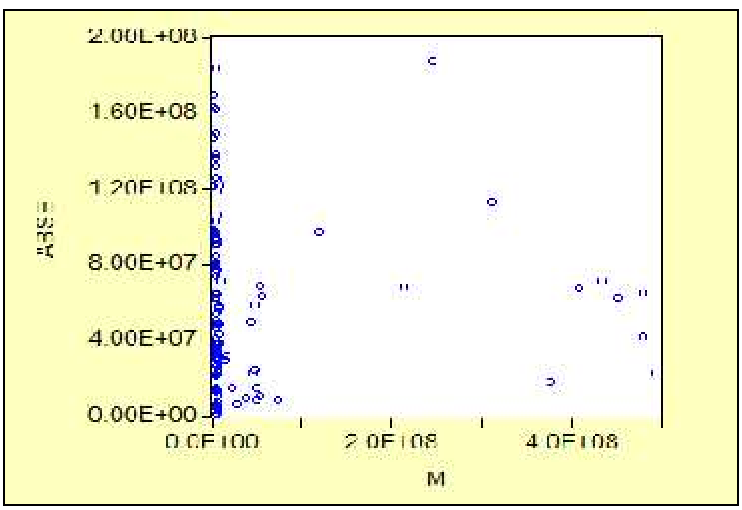
Cuadro 25. Diagrama de dispersión de valores absolutos de los errores frente al Consumo Privado (CP).



Como se puede observar en el Cuadro 25 conforme aumenta el valor del CP los errores se dispersan más, por lo que se puede concluir que existe heteroscedasticidad.

Finalmente se analizará gráficamente a la variable Importaciones (M) frente a los residuales en términos absolutos obteniendo el siguiente Cuadro:

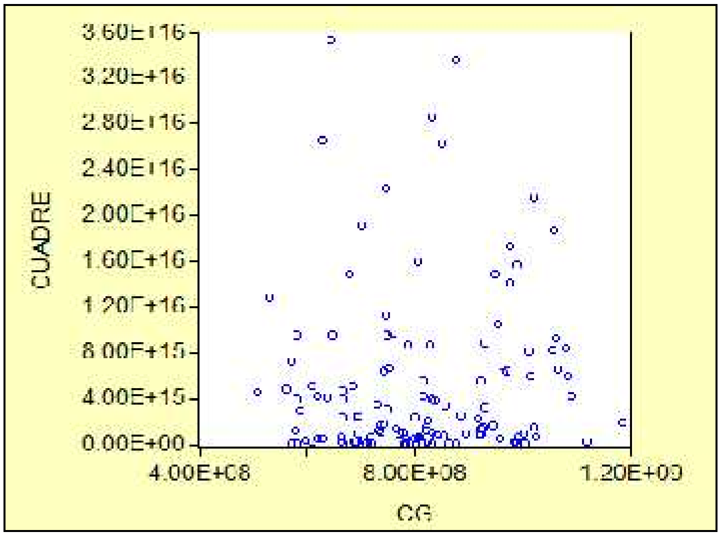
Cuadro 26. Diagrama de dispersión de valores absolutos de los errores frente a las Importaciones (M).



Como se hace evidente, a partir del Cuadro 26, los errores se concentran en su mayoría en los valores más pequeños de M, algunos pocos más se dispersan conforme aumenta el valor de M. Con este Cuadro se puede concluir que, debido a la alta concentración vertical de los errores, éstos son homoscedásticos.

Otro método gráfico para corroborar la conclusión de que existe heteroscedasticidad es el análisis de los errores al cuadrado frente a cada una de las variables independientes. La forma de obtención de dichos gráficos se explicó en el Ejemplo 1 de éste tema. Así, el diagrama de dispersión de los residuales al cuadrado frente al CG es el siguiente:

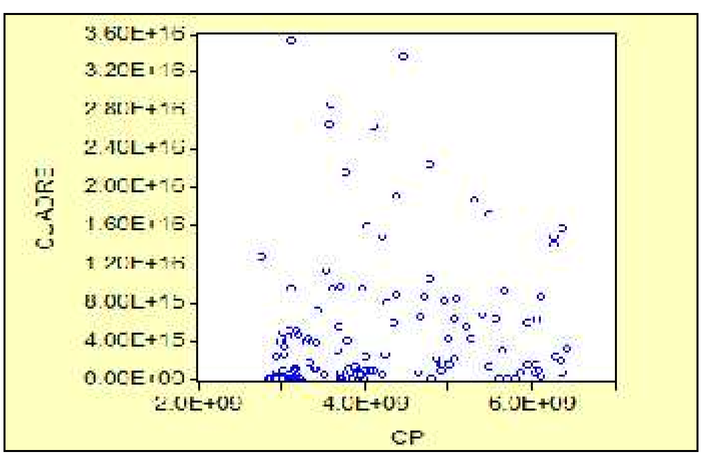
Cuadro 27. Diagrama de dispersión de los errores al cuadrado frente al Consumo Gubernamental (CG).



El análisis del Cuadro 27 es parecido a los anteriores, por lo que debido a que los valores de los errores elevados al cuadrado frente al CG presentan gran dispersión, es decir, no se concentran en torno a un mismo valor, se concluye que existe heteroscedasticidad. Asimismo, se observa que a medida que aumenta el valor del CG aumenta el valor del cuadrado de los errores, lo que confirma la presencia de heteroscedasticidad.

Para la variable CP se obtiene:

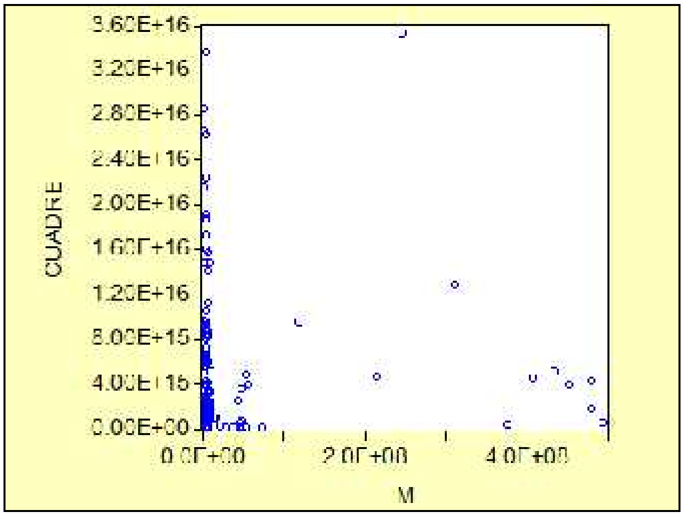
Cuadro 28. Diagrama de dispersión de los errores al cuadrado frente al Consumo Privado (CP).



En este caso se puede observar un resultado similar al del Cuadro anterior, ya que los valores de los errores al cuadrado frente al CP no se concentran en torno a un mismo valor. Asimismo, se observa que a medida que aumenta el valor del CP se incrementan los valores de los errores al cuadrado. Todo ello corrobora la presencia de heteroscedasticidad.

Ahora realizamos el diagrama de dispersión utilizando a la variable M:

Cuadro 29. Diagrama de dispersión de los errores al cuadrado frente a las Importaciones (M).



La interpretación del gráfico anterior es similar a la de los valores absolutos de los errores frente a las M, a saber: la mayor parte de los errores se concentran en los valores más pequeños de M, algunos más, están dispersos, sin embargo, la mayor concentración ocurre en los valores más pequeños de M, por lo que esta variable no es la causante de la heteroscedasticidad.

A continuación procedemos a realizar las pruebas formales para detectar la presencia de heteroscedasticidad en el modelo.

Método formal.

Prueba White para términos no cruzados.

Para la obtención de ésta prueba en EViews se realizó el mismo procedimiento que en la prueba del mismo nombre del ejemplo 1 de éste tema. Los resultados obtenidos son los siguientes:

Cuadro 30. Prueba White para términos no cruzados.

White Heteroskedasticity Test:

F-statistic	1.970283	Probability	0.075378
Obs*R-squared	11.37922	Probability	0.077340

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
White de términos no cruzados.
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.24E+16	2.30E+16	-1.407820	0.1618
CG	-23049054	47110107	-0.489259	0.6256
CG^2	0.013072	0.028030	0.466362	0.6418
CP	20310041	6878381	2.952736	0.0038
CP^2	0.002083	0.000736	2.833712	0.0054
M	85182430	34595573	2.462235	0.0153
M^2	-0.176553	0.073838	-2.391097	0.0184

R-squared	0.091768	Mean dependent var	5.23E+15
Adjusted R-squared	0.045192	S.D. dependent var	7.21E+15
S.E. of regression	7.05E+15	Akaike info criterion	75.87488
Sum squared resid	5.81E+33	Schwarz criterion	76.03409
Log likelihood	-4697.243	F-statistic	1.970283
Durbin-Watson stat	1.442727	Prob(F-statistic)	0.075378

Para esta prueba los valores más importantes, como ya se mencionó en el Ejemplo 1, son: el estadístico F y R^2.

Para hacer la interpretación del Cuadro 30 se plantean las siguientes hipótesis y reglas de decisión.

$$Ho: Homoscedasticidad.$$

$$Ho: \sigma_{ui}^2 = \sigma_{ui}^2$$

$$Ha: No\ Homoscedasticidad.$$

$$Ha: \sigma_{ui}^2 \neq \sigma_{ui}^2$$

El valor obtenido del estadístico F del Cuadro 30 se contrasta con F teórica, a un nivel de significación estadística de 5%, con 6 grados de libertad en el numerador y 117 en el denominador. Dicha determinación se realizó de la misma forma que se obtuvo en el ejemplo 1 de éste tema.

La regla de de decisión asociada a F estadística es la que se verifica a continuación:

$$Si\, F_c \leq F_\alpha \qquad Se\ acepta\ Ho$$
$$Si\, F_c > F_\alpha \qquad Se\ rechaza\ Ho$$

Entonces:

$$F_c = 1.970283$$
$$F_\alpha = 2.1769$$

Derivado de que:

$$F_c = 1.970283 < F_\alpha = 2.1769$$

Se acepta la Ho, es decir el modelo, de acuerdo a la prueba F, presenta homoscedasticidad.

Otra forma de detectar la presencia de heteroscedasticidad es mediante la probabilidad asociada al estadístico F, cuya regla de decisión es la siguiente:

$$Si\ Prob.\, asociada\ a\ F\ es < 0.05 \qquad Se\ rechaza\ Ho$$
$$Si\ Prob.\, asociada\ a\ F\ es \geq 0.05 \qquad Se\ acepta\ Ho$$

De acuerdo a lo anterior, el modelo presenta homoscedasticidad, ya que la Probabilidad asociada al estadístico F es igual a 0.075378 > 0.05, por lo que se acepta la hipótesis nula, es decir, el modelo presenta homoscedasticidad.

Por otra parte, el valor de R^2 que arroja el Cuadro 30 es de 0.091768, que también se puede expresar como 9.176%, lo que indica que las variables incluidas en la regresión auxiliar no explican el comportamiento, o lo hacen en una pequeña proporción, de los residuales elevados al cuadrado, concluyendo, a partir de esta prueba, que existe homoscedasticidad.

Prueba White para términos cruzados.

Para obtener esta prueba se realizaron los mismos pasos que en el ejemplo 1, obteniéndose los siguientes resultados:

Cuadro 31. Prueba White para términos cruzados.

White Heteroskedasticity Test:

F-statistic	1.295594	Probability	0.246926
Obs*R-squared	11.50628	Probability	0.242694

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
White para términos cruzados.
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.31E+16	2.45E+16	-1.349183	0.1800
CG	24910249	56779434	0.438720	0.6617
CG^2	0.022816	0.053194	0.428920	0.6688
CG*CP	0.003180	0.011312	0.281138	0.7791
CG*M	-0.017730	0.121159	-0.146330	0.8842
CP	20913857	7526865	2.778561	0.0064
CP^2	-0.001857	0.001059	-1.752502	0.0821
CP*M	-0.000166	0.054055	-0.003063	0.9976
M	96916574	1.22E+08	0.796296	0.4275
M^2	-0.175073	0.083811	-2.089520	0.0389

R-squared	0.092793	Mean dependent var	5.23E+15
Adjusted R-squared	0.021171	S.D. dependent var	7.21E+15
S.E. of regression	7.13E+15	Akaike info criterion	75.92214
Sum squared resid	5.80E+33	Schwarz criterion	76.14958
Log likelihood	-4697.173	F-statistic	1.295594
Durbin-Watson stat	1.442818	Prob(F-statistic)	0.246926

Para realizar la interpretación de los resultados mostrados en el Cuadro 31 se plantean las siguientes hipótesis y reglas de decisión.

$$Ho: Homoscedasticidad.$$
$$Ho: \sigma_{ui}^2 = \sigma_{ui}^2$$
$$Ha: No\ Homoscedasticidad.$$
$$Ha: \sigma_{ui}^2 \neq \sigma_{ui}^2$$

El valor obtenido del estadístico F en el Cuadro 31 se contrasta con F teórica, a un nivel de significación estadística de 5%, con 9 grados de libertad en el numerador y 114 en el denominador.

La regla de decisión asociada a F estadística es la que se muestra a continuación:

$$Si\ F_c \leq F_\alpha \qquad Se\ acepta\ Ho$$
$$Si\ F_c > F_\alpha \qquad Se\ rechaza\ Ho$$

Donde:

$$F_c = 1.295594$$
$$F_\alpha = 1.9629$$

Entonces:

$F_c = 1.295594 < F_\alpha = 1.9629$

Por lo que se acepta la Ho, es decir, el modelo presenta homoscedasticidad, de acuerdo al estadístico F.

Una manera más de detectar la presencia de heteroscedasticidad es mediante la probabilidad asociada al estadístico F, cuya regla de decisión es la siguiente:

$$\begin{array}{ll} Si\ Prob.\ asociada\ a\ F\ es < 0.05 & Se\ rechaza\ Ho \\ Si\ Prob.\ asociada\ a\ F\ es \geq 0.05 & Se\ acepta\ Ho \end{array}$$

Como se puede observar en el Cuadro 31 la Probabilidad asociada a F es 0.246926 > 0.05 por lo que se acepta la hipótesis nula, es decir, el modelo presenta homoscedasticidad. Los resultados de las pruebas White para términos cruzados y no cruzados conllevan a la misma conclusión, a saber: el modelo presenta homoscedasticidad. Este resultado contradice la conclusión obtenida mediante el análisis gráfico de los residuales, mismos que indicaban que el modelo presenta heteroscedasticidad.

Prueba ARCH.

Existe otra prueba para la detección de heteroscedasticidad, la prueba ARCH, la cual se obtiene mediante los comandos y pasos realizados en el ejemplo 1 de éste tema. En el Cuadro siguiente se muestran los resultados obtenidos:

Cuadro 32. Prueba ARCH LM.

ARCH Test:

F-statistic	14.34007	Probability	0.000239
Obs*R-squared	13.03257	Probability	0.000306

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Prueba ARCH LM
Sample (adjusted): 1980Q2 2010Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.46E+15	7.64E+14	4.522750	0.0000
RESID^2(-1)	0.324577	0.085712	3.786828	0.0002

R-squared	0.105956	Mean dependent var	5.17E+15
Adjusted R-squared	0.098567	S.D. dependent var	7.21E+15
S.E. of regression	6.84E+15	Akaike info criterion	75.77771
Sum squared resid	5.66E+33	Schwarz criterion	75.82343
Log likelihood	-4655.329	F-statistic	14.34007
Durbin-Watson stat	2.194759	Prob(F-statistic)	0.000239

Para la interpretación de los resultados mostrados en el Cuadro 32 se plantea la siguiente prueba de hipótesis:

Ho: Homoscedasticidad incondicional autorregresiva.
Ha: Heteroscedasticidad condicional autorregresiva de orden n.

Esta prueba está asociada también al estadístico F. El contraste del estadístico F se realiza de la misma forma que en la prueba de White, es decir, se contrasta F teórica con la F calculada, que es la que arroja el Cuadro 32. En cuanto a la F teórica se considero 1 grado de libertad en el numerador y 121 en el denominador, así como un nivel de significación del 5%.

Se conserva la regla de decisión que se utilizó en la prueba de White:

$Si\; F_c \le F_\alpha \qquad Se\ acepta\ Ho$

$Si\; F_c > F_\alpha \qquad Se\ rechaza\ Ho$

Donde:

$$F_\alpha = 3.9194$$
$$F_c = 14.34007$$

Entonces:

$F_c = 14.34007 > F_\alpha = 3.9194$, por tanto se rechaza la Ho, y se concluye con esta prueba que el modelo presenta problemas de heteroscedasticidad.

Como es evidente, para esta prueba de un rezago[21], y utilizando la probabilidad asociada a F y bajo la regla de decisión siguiente:

$Si\ Prob.\ asociada\ a\ F\ es < 0.05 \qquad Se\ rechaza\ Ho$

$Si\ Prob.\ asociada\ a\ F\ es \ge 0.05 \qquad Se\ acepta\ Ho$

Se concluye que la Probabilidad asociada a F es igual a 0.000239 < 0.05 es decir, se rechaza la hipótesis nula, lo que indica que el modelo presenta heteroscedasticidad condicional autorregresiva de orden 1. El resultado anterior no concuerda con el arrojado por las pruebas de White pero si con el mostrado por las pruebas gráficas. Por tanto, se procederá a solucionar dicho problema.

Corrección del modelo.

La corrección al modelo se realizó mediante la introducción de logaritmos, lo que convertirá al modelo en una forma funcional específica, a saber: nivel-logarítmica, es decir, las variables independientes se transformaron a logaritmos. Asimismo, dos de ellas, el CG y el CP, fueron rezagadas un periodo. Por lo que respecta a la variable dependiente, ésta continuará en su escala original, es decir, en niveles. Para introducir logaritmos al modelo se generan las variables tal y como se realizó en el ejemplo 1 del mismo tema.Para comprobar que el modelo presenta homoscedasticidad en su totalidad se realizarán las mismas pruebas formales que se realizaron con anterioridad.

[21] Obsérvese en el Cuadro 32 que la variable independiente es la variable dependiente, en este caso los residuos, pero rezagados un periodo.

Así, los resultados del nuevo modelo planteado son los siguientes:

Cuadro 33. Resultados del modelo corregido.

Dependent Variable: LUK
Method: Least Squares
Regresión Corregida.
Sample (adjusted): 1980Q2 2010Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.70E+10	1.20E+09	30.93373	0.0000
LCG(-1)	2.52E+08	66607626	3.788012	0.0002
LCP(-1)	1.90E+09	49080833	38.79245	0.0000
LM	77132956	7966022.	9.668196	0.0000

R-squared	0.949723	Mean dependent var	1.21E+09
Adjusted R-squared	0.940456	S.D. dependent var	4.16E+08
S.E. of regression	34405483	Akaike info criterion	39.59608
Sum squared resid	1.06E+18	Schwarz criterion	39.68753
Log likelihood	-2431.159	F-statistic	743.2983
Durbin-Watson stat	1.024959	Prob(F-statistic)	0.000000

El resultado de las pruebas formales se muestra a continuación:

1. **Prueba White para términos no cruzados.**

Se realiza esta prueba como se indicó con anterioridad con la finalidad de verificar la no existencia de heteroscedasticidad en el nuevo modelo propuesto. Los resultados de la prueba White para términos no cruzados se muestra a continuación:

Cuadro 34. Prueba White para términos no cruzados.

White Heteroskedasticity Test:

F-statistic	1.073532	Probability	0.382445
Obs*R-squared	6.470590	Probability	0.372586

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
White para términos no cruzados.
Sample: 1980Q2 2010Q4
Included observations: 123

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.39E+19	1.55E+19	0.895938	0.3721
LCG(-1)	-5.35E+17	1.19E+18	-0.451370	0.6526
LCG(-1)^2	1.32E+16	2.89E+16	0.455277	0.6498
LCP(-1)	-7.17E+17	1.12E+18	-0.642856	0.5216
LCP(-1)^2	1.60E+16	2.51E+16	0.636867	0.5255
LM	-4.44E+16	2.55E+16	-1.739503	0.0846
LM^2	1.23E+15	7.31E+14	1.676818	0.0963

R-squared	0.052606	Mean dependent var	8.62E+15
Adjusted R-squared	0.003603	S.D. dependent var	1.13E+16
S.E. of regression	1.13E+16	Akaike info criterion	76.81856
Sum squared resid	1.48E+34	Schwarz criterion	76.97861
Log likelihood	-4717.342	F-statistic	1.073532
Durbin-Watson stat	1.716864	Prob(F-statistic)	0.382445

Para realizar la interpretación de los resultados contenidos en el Cuadro 34 se plantea la siguiente prueba de hipótesis:

$H0: Homoscedasticidad.$

$Ho: \sigma_{ui}^2 = \sigma_{ui}^2$

$Ha: No\ Homoscedasticidad.$

$Ha: \sigma_{ui}^2 \neq \sigma_{ui}^2$

El valor del estadístico F mostrado en el Cuadro 31 se contrasta con F teórica, a un nivel de significación estadística de 5%, con 6 grados de libertad en el numerador y 116 en el denominador.
La regla de decisión es la que se indica a continuación:

$Si\ F_c \leq F_\alpha \quad Se\ acepta\ Ho$

$Si\ F_c > F_\alpha \quad Se\ rechaza\ Ho$

Entonces:

$F_c = 1.073532$

$F_\alpha = 2.1776$

Por lo tanto, se observa que:

$$F_c = 1.073532 < F_\alpha = 2.1776$$

De acuerdo a este resultado se acepta la Ho, concluyendo que el modelo presenta homoscedasticidad.

Otra forma de detectar la presencia o no de la heteroscedasticidad relacionada con el estadístico F es mediante su probabilidad asociada, cuya regla de decisión es la siguiente:

$Si\ Prob.\ asociada\ a\ F\ es < 0.05 \quad Se\ rechaza\ Ho$

$Si\ Prob.\ asociada\ a\ F\ es \geq 0.05 \quad Se\ acepta\ Ho$

En base a lo anterior se puede concluir que el modelo presenta homoscedasticidad ya que la Probabilidad asociada a F es 0.382445 > 0.05, por lo quese acepta la hipótesis nula, el modelo presenta homoscedasticidad.

Ello también se corrobora con el valor que asume el R^2, mismo que asciende a 0.052606, que también se puede expresar como porcentaje, 5.26%, indicando que las variables explicativas, en realidad están explicando muy poco a los residuales elevados al cuadrado.

Prueba White para términos cruzados.

Siguiendo los pasos descritos en el ejemplo 1 de este tema, pero aplicándolos al caso concreto de este modelo, se obtuvieron los siguientes resultados:

Cuadro 35. Prueba White para términos cruzados.

White Heteroskedasticity Test:

F-statistic	1.620897	Probability	0.117511
Obs*R-squared	14.06349	Probability	0.120085

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
White de términos cruzados.
Sample: 1980Q2 2010Q4
Included observations: 123

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5.03E+18	1.79E+19	-0.280841	0.7793
LCG(-1)	1.75E+18	1.53E+18	1.147172	0.2537
LCG(-1)^2	-1.15E+17	5.95E+16	-1.929195	0.0562
LCG(-1)*LCP(-1)	1.43E+17	6.90E+16	2.067934	0.0409
LCG(-1)*LM	-1.26E+16	9.48E+15	-1.324066	0.1882
LCP(-1)	-1.44E+18	1.17E+18	-1.231141	0.2208
LCP(-1)^2	-3.09E+16	3.41E+16	-0.904849	0.3675
LCP(-1)*LM	-7.62E+15	1.22E+16	-0.625721	0.5328
LM	4.06E+17	2.55E+17	1.592510	0.1141
LM^2	3.86E+14	8.50E+14	0.454161	0.6506

R-squared	0.114337	Mean dependent var	8.62E+15
Adjusted R-squared	0.043798	S.D. dependent var	1.13E+16
S.E. of regression	1.11E+16	Akaike info criterion	76.79996
Sum squared resid	1.38E+34	Schwarz criterion	77.02860
Log likelihood	-4713.198	F-statistic	1.620897
Durbin-Watson stat	1.734736	Prob(F-statistic)	0.117511

Para realizar la interpretación de los resultados mostrados en el Cuadro 35 se recurre a la misma prueba de hipótesis y reglas de decisión que se utilizaron en la prueba anterior. Así:

$$H0: Homoscedasticidad.$$

$$Ho: \sigma_{ui}^2 = \sigma_{ui}^2$$

$$Ha: No\ Homoscedasticidad.$$

$$Ha: \sigma_{ui}^2 \neq \sigma_{ui}^2$$

Para contrastar el valor del estadístico F obtenido en el Cuadro 35 con F teórica, a un nivel de significación estadística de 5%, con 9 grados de libertad en el numerador y 113 en el denominador, es necesario valerse de la siguiente regla de decisión.

$$Si\ F_c \leq F_\alpha \qquad Se\ acepta\ Ho$$
$$Si\ F_c > F_\alpha \qquad Se\ rechaza\ Ho$$

Donde:

$$F_c = 1.620897$$
$$F_\alpha = 1.9637$$

Entonces:

$$F_c = 1.620897 < F_\alpha = 1.9637$$

Por tanto se acepta la Ho, por lo que el modelo presenta homoscedasticidad.

Una manera adicional de detectar la presencia de heteroscedasticidad es mediante la probabilidad asociada al estadístico F, cuya regla de decisión es la siguiente:

$$Si\ Prob.asociada\ a\ F\ es < 0.05 \qquad Se\ rechaza\ Ho$$
$$Si\ Prob.asociada\ a\ F\ es \geq 0.05 \qquad Se\ acepta\ Ho$$

En el Cuadro 35 la Probabilidad asociada a F es 0.117511 > 0.05 por lo que se acepta la hipótesis nula, es decir, el modelo presenta homoscedasticidad. Los resultados de ambas pruebas de White detectan la presencia de homoscedasticidad.

Ahora procederemos a realizar la prueba ARCH para determinar si en efecto el modelo ya no presenta heteroscedasticidad.

Prueba ARCH.

Esta prueba se realiza utilizando el mismo mecanismo empleado para su obtención en el ejemplo 1 del mismo tema.

Cuadro 36. Prueba ARCH LM.

ARCH Test:

F-statistic	3.161383	Probability	0.077932
Obs*R-squared	3.131572	Probability	0.076790

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Prueba ARCH
Sample (adjusted): 1980Q3 2010Q4
Included observations: 122 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.19E+15	1.29E+15	5.594638	0.0000
RESID^2(-1)	0.160468	0.090250	1.778028	0.0779

R-squared	0.025669	Mean dependent var	8.59E+15
Adjusted R-squared	0.017549	S.D. dependent var	1.14E+16
S.E. of regression	1.13E+16	Akaike info criterion	76.77248
Sum squared resid	1.52E+34	Schwarz criterion	76.81845
Log likelihood	-4681.121	F-statistic	3.161383
Durbin-Watson stat	1.969530	Prob(F-statistic)	0.077932

Se utiliza la misma prueba de hipótesis y regla de decisión que en la regresión original.

Ho: Homoscedasticidad incondicional autorregresiva.
Ha: Heteroscedasticidad condicional autorregresiva de orden n.

Para determinar la presencia de homoscedasticidad o heteroscedasticidad en el modelo a partir de esta prueba, al igual que en las pruebas de White, se contrasta el estadístico F calculado con la F teórica. Para el caso de esta última se considera 1 grado de libertad en el numerador y 120 en el denominador, así como un nivel de significación de 5%.

Así, se utiliza la misma regla de decisión que se utilizó en la prueba de White:

$$Si\ F_c \leq F_\alpha \qquad Se\ acepta\ Ho$$
$$Si\ F_c > F_\alpha \qquad Se\ rechaza\ Ho$$

Donde:

$$F_\alpha = 3.9201$$
$$F_c = 3.161383$$

Por lo tanto:

$F_c = 3.161383 < F_\alpha = 3.9201$ indicando que se acepta la Ho, y por tanto, se concluye que, en base al resultado de esta prueba, el modelo presenta homoscedasticidad.

Continuando con la evaluación de los valores de F, se procede a contrastar la probabilidad asociada a F, bajo la siguiente regla de decisión:

$$\begin{array}{ll} Si\ Prob.\ asociada\ a\ F\ es\ < 0.05 & Se\ rechaza\ Ho \\ Si\ Prob.\ asociada\ a\ F\ es\ \geq 0.05 & Se\ acepta\ Ho \end{array}$$

Como la Probabilidad asociada a F es igual a 0.077932 >0.05, se acepta la hipótesis nula, es decir, el modelo presenta homoscedasticidad incondicional autorregresiva.

Por tanto, concluimos que el nuevo modelo planteado no presenta problemas de heteroscedasticidad de acuerdo con los resultados arrojados por cada una de las pruebas realizadas con anterioridad.

Ejemplo 3. Producto Interno Bruto en función del CP, X y FBK.

Para realizar este ejemplo se utilizará como variable dependiente el Producto Interno Bruto (PIB) a precios de 2003 y como variables independientes el Consumo Privado (CP), las Exportaciones (X) en miles de pesos y la Formación Bruta de Capital (FBK). El periodo de los datos comprende desde el primer trimestre de 1980 hasta el cuarto trimestre de 2010.

Así, comenzaremos llevando a cabo la regresión siguiente:

PIB= f (CP, X, FBK)

Para llevar a cabo dicha regresión con EViews se seguirán los pasos ya habitúales, obteniendo el siguiente Cuadro:

Cuadro 37. Regresión original.

Dependent Variable: PIB
Method: Least Squares

Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.30E+09	2.80E+08	4.650990	0.0000
CP	1.143623	0.106016	10.83448	0.0000
X	0.569483	0.268920	2.117665	0.0363
FBK	0.062955	0.206033	0.305556	0.7605

R-squared	0.971275	Mean dependent var	6.40E+09
Adjusted R-squared	0.970557	S.D. dependent var	1.49E+09
S.E. of regression	2.55E+08	Akaike info criterion	41.58359
Sum squared resid	7.81E+18	Schwarz criterion	41.67457
Log likelihood	-2574.183	F-statistic	1352.524
Durbin-Watson stat	0.540024	Prob(F-statistic)	0.000000

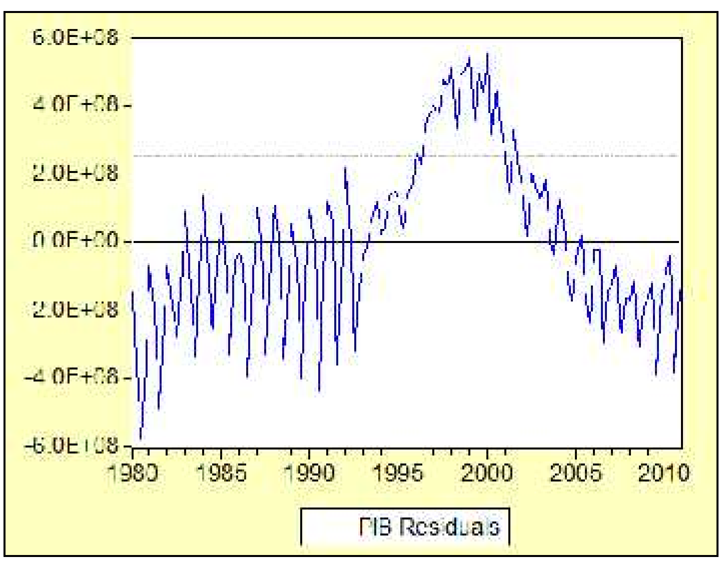
Con los resultados obtenidos en el Cuadro 37 se puede realizar el análisis de detección de heteroscedasticidad a partir de los valores que arrojan S.E. of regression y Sum squared resid, que deberían de ser bajos en presencia de homoscedasticidad. El valor de S.E. of regression es $2.55x10^{8}$ y Sum squared resid es de $7.81x10^{18}$, los cuales son valores muy altos, lo que podría, a su vez, indicar que existe heteroscedasticidad.

Método gráfico.

Gráfica de residuos.

A continuación se presenta la gráfica de los residuos frente al tiempo:

Cuadro 38. Residuos frente al tiempo.



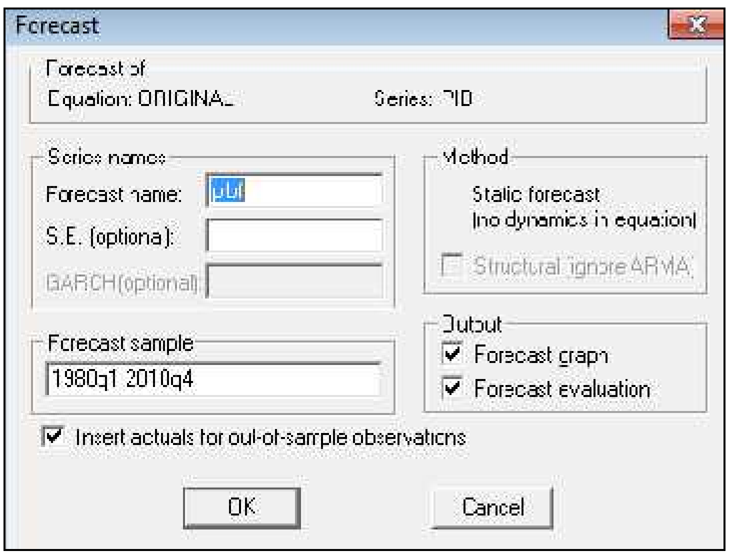
En este Cuadro se observa que los errores del PIB frente al tiempo presentan una heteroscedasticidad severa, ya que la serie se sale de las bandas de confianza[22].

Diagrama de dispersión entre los valores absolutos de los errores, el cuadrado de los mismos y la variable independiente.

Para corroborar la presencia de heteroscedasticidad en el modelo se realizan los diagramas de dispersión de los errores en valores absolutos y elevados al cuadrado frente a los valores estimados del PIB.

Para obtener los valores estimados del PIB $(P\hat{I}B)$ en Eviews 5, una vez que se ha realizado la regresión (ver Cuadro 37) en dicha ventana vamos al menú y damos un click sobre el botón “Forecast”, con lo cual se desplegará una ventana como la siguiente:

Cuadro 39. Estimación.

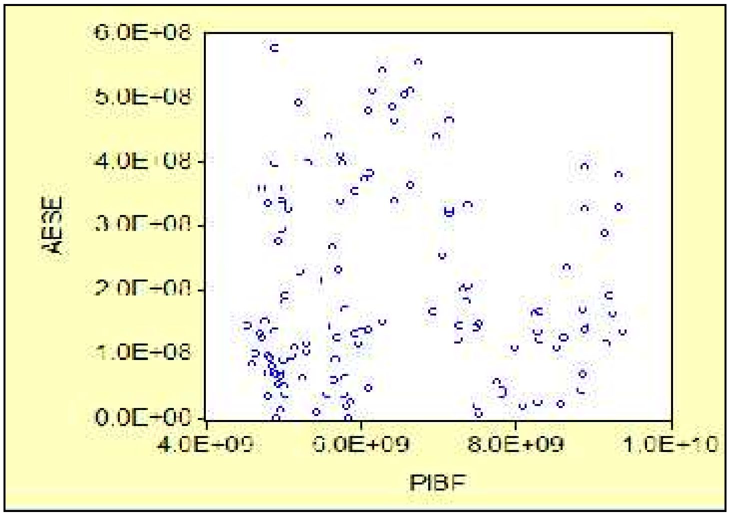


En este caso observamos que el nombre de la serie estimada es “PIBF” y que el periodo de estimación comprende desde primer trimestre de 1980 hasta el cuarto de 2010. Seleccionamos la opción “Ok” de dicho Cuadro y ahora la estimación del PIB (PIBF), realizada a partir de la ecuación de regresión obtenida en el Cuadro 37, aparecerá en la ventana del workfile.

A continuación, para realizar el diagrama de dispersión procedemos a generar los valores absolutos de los errores, “ABSE”, como se indicó en el ejemplo 1 de este tema. Siguiendo los mismos pasos descritos anteriormente para obtener el diagrama de dispersión entre PIBF y ABSE obtenemos el siguiente Cuadro:

[22] Éstas dan el rango dentro del cual dicha serie puede fluctuar sin presentar heteroscedasticidad.

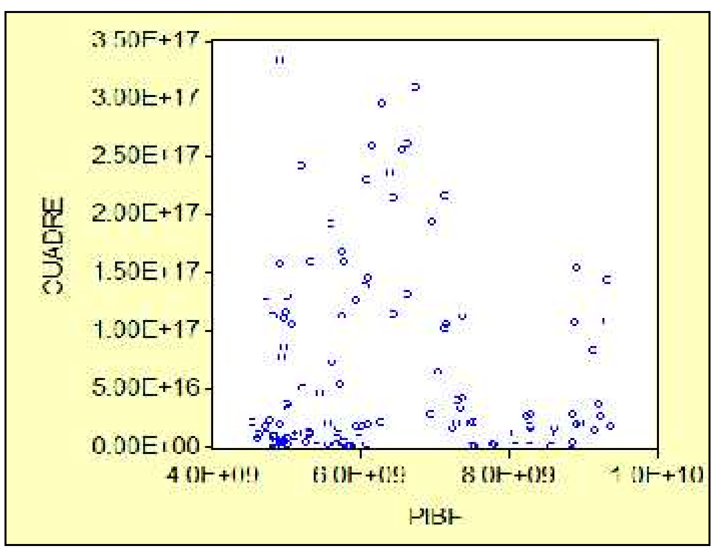
Cuadro 40. Diagrama de dispersión del valor absoluto de los errores frente al PIB estimado.



En el Cuadro anterior se puede observar que a medida que se incrementan los valores del PIBF también lo hacen los valores de los errores en términos absolutos, mostrando, por tanto, gran dispersión. Ello es muestra de que el modelo presenta heteroscedasticidad.

Ahora, para obtener el diagrama de dispersión de los valores al cuadrado de los errores frete al PIB estimado se llevará a cabo el mismo procedimiento utilizado para obtener el gráfico anterior, con excepción de que ahora se deberá crear la serie "CUADRE" como se indicó anteriormente. Así, el resultado es el siguiente:

Cuadro 40. Diagrama de dispersión de los errores al cuadrado frente al PIB estimado.



En este diagrama de dispersión se observa que si bien hay una mayor concentración de los valores, a medida que se incrementan los valores del PIB estimado también crecen los de los errores al cuadrado, denotando heteroscedasticidad en el modelo.

De acuerdo con las pruebas gráficas se puede inferir que el modelo en general presenta heteroscedasticidad, sin embargo, es necesario realizar otro tipo de pruebas, en este caso formales, para corroborar que la afirmación que se está haciendo es correcta.

Método formal.

Prueba White para términos no cruzados.

La obtención de ésta prueba se realizó con base en el mismo procedimiento usado en los ejemplos 1 y 2 del tema de heteroscedasticidad. Así, siguiendo los mismos pasos se obtiene:

Cuadro 41. Prueba White para términos no cruzados.

White Heteroskedasticity Test:

F-statistic	5.175500	Probability	0.000094
Obs*R-squared	26.00306	Probability	0.000222

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Date: 10/23/11 Time: 19:57
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.04E+16	2.27E+17	-0.265796	0.7909
CP	84729579	1.42E+08	0.594844	0.5531
CP^2	-0.025906	0.017417	-1.487335	0.1396
X	3.32E+08	1.29E+08	2.572257	0.0114
X^2	-0.194801	0.116132	-1.677402	0.0961
FBK	52454735	2.12E+08	0.247193	0.8052
FBK^2	0.080895	0.096313	0.839920	0.4027

R-squared	0.209742	Mean dependent var	6.30E+16
Adjusted R-squared	0.169216	S.D. dependent var	8.03E+16
S.E. of regression	7.32E+16	Akaike info criterion	80.55734
Sum squared resid	6.27E+35	Schwarz criterion	80.71655
Log likelihood	-4987.555	F-statistic	5.175500
Durbin-Watson stat	1.644358	Prob(F-statistic)	0.000094

Con el fin de analizar los resultados contenidos en el Cuadro 41 se plantea la siguiente prueba de hipótesis:

$$H0: Homoscedasticidad.$$

$$Ho: \sigma_{ui}^2 = \sigma_{ui}^2$$
$$Ha: No\ Homoscedasticidad.$$
$$Ha: \sigma_{ui}^2 \neq \sigma_{ui}^2$$

El valor del estadístico F arrojado en el Cuadro 41 se contrasta con F teórica, a un nivel de significación estadística de 5%, con 6 grado de libertad en el numerador y 117 en el denominador. Esta determinación se realizó de la misma forma que se obtuvo en el ejemplo 1 de éste tema.

La regla de de decisión asociada al estadístico F es la que se muestra a continuación:

$Si\ F_c \leq F_\alpha \qquad Se\ acepta\ Ho$

$Si\ F_c > F_\alpha \qquad Se\ rechaza\ Ho$

Entonces:

$F_c = 5.1755$

$F_\alpha = 2.1769$

Se observa que:

$$F_c = 5.1755 > F_\alpha = 2.1769$$

Por lo que se puede concluir que el modelo presenta heteroscedasticidad, aceptando la Ha.

Existe otra forma de detectar la presencia de heteroscedasticidad con la prueba F y es mediante la probabilidad asociada, cuya regla de decisión es la siguiente:

$Si\ Prob.asociada\ a\ F\ es < 0.05 \qquad Se\ rechaza\ Ho$

$Si\ Prob.asociada\ a\ F\ es \geq 0.05 \qquad Se\ acepta\ Ho$

La Probabilidad arrojada por el Cuadro 41 es de 0.000094 < 0.05 por lo que se rechaza la Ho, es decir, el modelo presenta heteroscedasticidad.

Asimismo, el valor de R^2 es de 0.2097, que en porcentaje es 20.97%, indicando que las variables independientes de la regresión auxiliar (prueba White) están explicando a la variable dependiente, lo cual no debería ocurrir.

Prueba White para términos cruzados.

La obtención de esta prueba se realiza conforme a lo establecido para obtener la prueba del mismo nombre del ejemplo 1 dentro del tema de heteroscedasticidad. Se obtuvieron los siguientes resultados:

Cuadro 42. Prueba de White para términos cruzados.

White Heteroskedasticity Test:

F-statistic	4.321384	Probability	0.000074
Obs*R-squared	31.54285	Probability	0.000239

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.16E+18	9.87E+17	-2.191864	0.0304
CP	1.50E+09	6.48E+08	2.317286	0.0223
CP^2	-0.315445	0.132136	-2.387275	0.0186
CP*X	1.185264	0.541668	2.188176	0.0307
CP*FBK	0.535642	0.412408	1.298817	0.1966
X	-3.42E+09	1.75E+09	-1.958237	0.0526
X^2	-1.688249	0.765170	-2.206369	0.0294
X*FBK	-0.323794	1.104146	-0.293253	0.7699
FBK	-4.38E+08	1.05E+09	-0.419096	0.6759
FBK^2	-0.584067	0.430629	-1.356311	0.1777

R-squared	0.254378	Mean dependent var	6.30E+16
Adjusted R-squared	0.195513	S.D. dependent var	8.03E+16
S.E. of regression	7.21E+16	Akaike info criterion	80.54759
Sum squared resid	5.92E+35	Schwarz criterion	80.77503
Log likelihood	-4983.951	F-statistic	4.321384
Durbin-Watson stat	1.748330	Prob(F-statistic)	0.000074

Con el fin de analizar los resultados mostrados en el Cuadro 42 se establece la siguiente prueba de hipótesis:

$$H0: Homoscedasticidad.$$

$$Ho: \sigma_{ui}^2 = \sigma_{ui}^2$$

$$Ha: No\ Homoscedasticidad.$$

$$Ha: \sigma_{ui}^2 \neq \sigma_{ui}^2$$

La regla de decisión es la siguiente:

$$Si\ F_c \leq F_\alpha \qquad Se\ acepta\ Ho$$
$$Si\ F_c > F_\alpha \qquad Se\ rechaza\ Ho$$

Para determinar el valor de la F teórica se determina que los grados de libertad en el numerador son de 9 y los grados de libertad en el denominador de 114, utilizando un nivel de significancia de 5%, F teórica es de 1.9629

Mediante la regla de decisión se observa que F_c= 4.3213 >$F_{\alpha} = 1.9629$ por lo que se procede a rechazar la hipótesis nula, es decir, el modelo presenta heteroscedasticidad.

Otra forma de detección es por medio de la probabilidad asociada a F, bajo la siguiente regla de decisión.

$$Si\ Prob.\ asociada\ a\ F\ es\ < 0.05 \qquad Se\ rechaza\ Ho$$
$$Si\ Prob.\ asociada\ a\ F\ es\ \geq 0.05 \qquad Se\ acepta\ Ho$$

La Probabilidad asociada al estadístico F es de 0.000074 < 0.05 por lo que se rechaza la hipótesis nula, es decir, el modelo presenta heteroscedasticidad.

Prueba ARCH.

Otro método para detectar la presencia de heteroscedasticidad en el modelo es la prueba de Heteroscedasticidad Condicional Autorregresiva (ARCH).

Para generar el Cuadro concerniente a ésta prueba se siguieron los mismos pasos que en el ejemplo 1. Los resultados son los siguientes:

Cuadro 43. Prueba ARCH.

ARCH Test:

F-statistic	30.72592	Probability	0.000000
Obs*R-squared	24.90865	Probability	0.000001

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample (adjusted): 1980Q2 2010Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.48E+16	8.30E+15	4.192184	0.0001
RESID^2(-1)	0.450095	0.081199	5.543097	0.0000

R-squared	0.202509	Mean dependent var	6.33E+16
Adjusted R-squared	0.195919	S.D. dependent var	8.06E+16
S.E. of regression	7.23E+16	Akaike info criterion	80.49196
Sum squared resid	6.32E+35	Schwarz criterion	80.53769
Log likelihood	-4948.256	F-statistic	30.72592
Durbin-Watson stat	2.211231	Prob(F-statistic)	0.000000

Con el fin de analizar los resultados obtenidos en el Cuadro 43 se establece la siguiente prueba de hipótesis:

Ho: Homoscedasticidad incondicional autorregresiva.
Ha: Heteroscedasticidad condicional autorregresiva de orden n.

Para determinar la posible existencia de heteroscedasticidad en el modelo se utiliza el estadístico F. Para ello se establece la siguiente regla de decisión:

$$\begin{array}{ll} Si\, F_c \leq F_\alpha & Se\ acepta\ Ho \\ Si\, F_c > F_\alpha & Se\ rechaza\ Ho \end{array}$$

Por lo que respecta al valor de F teórica se utiliza 1 grado de libertad en el numerador y 121 en el denominador, utilizando un nivel de significación de 5%:

$$F_\alpha = 3.9194$$

Entonces:

$F_c = 30.7259 > F_\alpha = 3.9194$ lo que indica que el modelo presenta heteroscedasticidad.

Dicha conclusión se corrobora mediante el análisis de la probabilidad asociada al estadístico F. La regla de decisión es la siguiente.

$$\begin{array}{ll} Si\ Prob.\ asociada\ a\ F\ es < 0.05 & Se\ rechaza\ Ho \\ Si\ Prob.\ asociada\ a\ F\ es \geq 0.05 & Se\ acepta\ Ho \end{array}$$

La Probabilidad obtenida es de 0.0000 < 0.05, por lo tanto se rechaza la Ho, es decir, el modelo presenta heteroscedasticidad.

En conclusión, y de acuerdo con las pruebas realizadas, el modelo presenta problemas de heteroscedasticidad, tal y como se planteó desde un principio, a continuación se procederá a corregir este problema.

Corrección del modelo.

Para corregir el problema de heteroscedasticidad el modelo ha quedado definido de la siguiente manera: se modificó la forma funcional quedando un modelo doble logarítmico, es decir, tanto la variable explicada como las explicativas fueron transformadas a logaritmos naturales. Para introducir logaritmos en este modelo se ocuparon los mismos pasos que en la introducción de logaritmos del ejemplo 1 de este tema. El Consumo Privado fue rezagado en un periodo, las Exportaciones en dos, mientras que la Formación Bruta de Capital se rezago en dos periodos.

Los resultados del nuevo modelo planteado son los siguientes:

Cuadro 44. Regresión corregida.

Dependent Variable: LPIB
Method: Least Squares

Sample (adjusted): 1980Q3 2010Q4
Included observations: 122 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	9.788493	0.739685	13.23332	0.0000
LCP(-1)	0.356414	0.059606	5.979542	0.0000
LX(-2)	0.030558	0.002826	10.81117	0.0000
LFBK(-2)	0.207986	0.031655	6.570443	0.0000

R-squared	0.973652	Mean dependent var	22.55841
Adjusted R-squared	0.972982	S.D. dependent var	0.223470
S.E. of regression	0.037554	Akaike info criterion	-3.693853
Sum squared resid	0.166413	Schwarz criterion	-3.601918
Log likelihood	229.3250	F-statistic	1453.519
Durbin-Watson stat	1.478390	Prob(F-statistic)	0.000000

En esta regresión se analiza, al igual que en la regresión original, el valor del S.E. of regression, el cual es de 0.037554 y de Sum squared resid cuyo valor es de 0.166413; ambos son valores ínfimos si se comparan con los obtenidos en la regresión original, lo cual puede ser un signo de que le heteroscedasticidad se ha corregido, sin embargo, como ya se ha mencionado, es necesario aplicar otras pruebas para corroborar lo anterior.

1. **Prueba White para términos no cruzados.**

Esta prueba se obtuvo aplicando el mismo procedimiento que en los ejercicios anteriores. El resultado se muestra a continuación:

Cuadro 45. Prueba de White para términos no cruzados.

White Heteroskedasticity Test:

F-statistic	1.292840	Probability	0.266024
Obs*R-squared	7.709200	Probability	0.260191

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample: 1980Q3 2010Q4
Included observations: 122

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.186041	3.015532	-0.393311	0.6948
LCP(-1)	0.353580	0.373242	0.947322	0.3455
LCP(-1)^2	-0.008021	0.008403	-0.954524	0.3418
LX(-2)	0.003257	0.001990	1.636251	0.1045
LX(-2)^2	-0.000103	6.28E-05	-1.641401	0.1034
LFBK(-2)	-0.265200	0.151188	-1.754111	0.0821
LFBK(-2)^2	0.006430	0.003639	1.766882	0.0799

R-squared	0.063190	Mean dependent var	0.001364
Adjusted R-squared	0.014313	S.D. dependent var	0.002131
S.E. of regression	0.002115	Akaike info criterion	-9.423425
Sum squared resid	0.000515	Schwarz criterion	-9.262538
Log likelihood	581.8289	F-statistic	1.292840
Durbin-Watson stat	1.932737	Prob(F-statistic)	0.266024

Para analizar los resultados obtenidos se plantea la siguiente prueba de hipótesis:

$$H0: Homoscedasticidad.$$
$$Ho: \sigma_{ui}^2 = \sigma_{ui}^2$$
$$Ha: No\ Homoscedasticidad.$$
$$Ha: \sigma_{ui}^2 \neq \sigma_{ui}^2$$

Para realizar el análisis se contrasta el valor del estadístico F obtenido en el Cuadro 39 contra la F teórica. Para determinar el valor de esta última se considera un nivel de significación estadística de 5%, con 6 grados de libertad en el numerador y 115 en el denominador.

La regla de de decisión es la siguiente:

$$Si\ F_c \leq F_\alpha \quad Se\ acepta\ Ho$$
$$Si\ F_c > F_\alpha \quad Se\ rechaza\ Ho$$

Por lo que:

$F_c = 1.309898$
$F_\alpha = 2.1783$

Se verifica que:

$$F_c = 1.309898 < F_\alpha = 2.1783$$

Por tanto, se acepta la hipótesis nula, es decir, el modelo es homoscedástico.

Otra forma de detectar la presencia de heteroscedasticidad en el modelo con la prueba F es mediante su probabilidad asociada, cuya regla de decisión es la siguiente:

$Si\ Prob.\ asociada\ a\ F\ es < 0.05 \qquad Se\ rechaza\ Ho$
$Si\ Prob.\ asociada\ a\ F\ es \geq 0.05 \qquad Se\ acepta\ Ho$

La Probabilidad del estadístico F es de 0.2660 > 0.05 por lo que se acepta la Ho, es decir, el modelo presenta homoscedasticidad. El valor obtenido de R^2 del mismo Cuadro es de 0.0631, que en porcentaje se expresa como, 6.31%, indicando que las variables explicativas, están explicando poco a la variable dependiente. Concluyéndose que el modelo no presenta heteroscedasticidad.

Prueba White para términos cruzados.

A continuación se muestran los resultados obtenidos con la prueba White para términos cruzados. Cabe comentar que dicha prueba se realizó siguiendo el mismo procedimiento aplicado en los dos ejemplos anteriores.

Cuadro 46. Prueba de White para términos cruzados.

White Heteroskedasticity Test:

F-statistic	1.307387	Probability	0.246937
Obs*R-squared	10.33549	Probability	0.242264

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample: 1980Q3 2010Q4
Included observations: 122

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.968263	5.884410	1.014250	0.3126
LCP(-1)	-0.600430	0.575813	-1.042751	0.2993
LCP(-1)^2	0.029007	0.014714	1.971344	0.0511
LCP(-1)*LX(-2)	-0.004366	0.003441	-1.268582	0.2072
LCP(-1)*LFBK(-2)	-0.028993	0.017308	-1.675144	0.0967
LX(-2)	0.072158	0.059657	1.209553	0.2290
LX(-2)^2	2.07E-06	0.000127	0.016307	0.9870
LX(-2)*LFBK(-2)	0.001139	0.000834	1.365740	0.1747
LFBK(-2)^2	0.014964	0.009251	1.617605	0.1085

R-squared	0.084717	Mean dependent var	0.001364
Adjusted R-squared	0.019918	S.D. dependent var	0.002131
S.E. of regression	0.002109	Akaike info criterion	-9.413885
Sum squared resid	0.000503	Schwarz criterion	-9.207031
Log likelihood	583.2470	F-statistic	1.307387
Durbin-Watson stat	1.959002	Prob(F-statistic)	0.246937

Con el fin de analizar los resultados contenidos en el Cuadro 46 se plantea la siguiente prueba de hipótesis:

$$HO: Homoscedasticidad.$$
$$Ho: \sigma_{ui}^2 = \sigma_{ui}^2$$
$$Ha: No\ Homoscedasticidad.$$
$$Ha: \sigma_{ui}^2 \neq \sigma_{ui}^2$$

La regla de decisión es la siguiente:

$Si\ F_c \leq F_\alpha \quad Se\ acepta\ Ho$

$Si\ F_c > F_\alpha \quad Se\ rechaza\ Ho$

El valor de la F teórica se determina utilizando 8 grados de libertad para el numerador y 113 para el denominador, asimismo, el valor del nivel de significancia es del 5%. Así, el valor de ésta es de:

$$F_\alpha = 2.0213$$

Dado que F_c= 1.3073 < $F_{\alpha} = 2.0213$ se acepta la hipótesis nula, es decir, el modelo es homoscedástico.

La probabilidad asociada a F es otra forma de detectar la existencia de heteroscedasticidad, se contrasta bajo la siguiente regla de decisión.

$$\begin{array}{ll} Si\ Prob.\ asociada\ a\ F\ es < 0.05 & Se\ rechaza\ Ho \\ Si\ Prob.\ asociada\ a\ F\ es \geq 0.05 & Se\ acepta\ Ho \end{array}$$

En este caso la Probabilidad asociada a F es de 0.2469 > 0.05, por lo que se acepta la hipótesis nula, es decir, el modelo presenta homoscedasticidad.

Asimismo, el valor que asume R^2 = 0.0847 = 8.47%, indica que las variables independientes incluidas en la regresión auxiliar (prueba White) no explican el comportamiento de los residuales al cuadrado, por lo que se corrobora que el modelo no presenta heteroscedasticidad.

Prueba ARCH.

Finalmente se lleva a cabo la prueba ARCH para determinar la presencia o ausencia de heteroscedasticidad en el modelo. Dicha prueba se realizó siguiendo el procedimiento establecido en el ejemplo 1 de este tema. Los resultados son los siguientes:

Cuadro 47. Prueba ARCH.

ARCH Test:

F-statistic	1.120176	Probability	0.292024
Obs*R-squared	1.128380	Probability	0.288121

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares

Sample (adjusted): 1980Q4 2010Q4
Included observations: 121 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001233	0.000231	5.346765	0.0000
RESID^2(-1)	0.096534	0.091256	1.058383	0.2920

R-squared	0.009325	Mean dependent var	0.001365
Adjusted R-squared	0.001000	S.D. dependent var	0.002140
S.E. of regression	0.002139	Akaike info criterion	-9.440999
Sum squared resid	0.000544	Schwarz criterion	-9.394788
Log likelihood	573.1805	F-statistic	1.120176
Durbin-Watson stat	1.992630	Prob(F-statistic)	0.292024

Para llevar a cabo el análisis de los resultados obtenidos se plantea la siguiente prueba de hipótesis:

Ho: Homoscedasticidad incondicional autorregresiva.
Ha: Heteroscedasticidad condicional autorregresiva de orden n.

Para determinar la presencia de homoscedasticidad o heteroscedasticidad en el modelo en base a esta prueba, al igual que en las pruebas de White, se contrasta el estadístico F calculado con la F teórica. Para el caso de esta última se considera 1 grado de libertad en el numerador y 119 en el denominador, así como un nivel de significación de 5%.

Así, se utiliza la misma regla de decisión que se utilizó en la prueba de White:

$Si\ F_c \leq F_\alpha \qquad Se\ acepta\ Ho$
$Si\ F_c > F_\alpha \qquad Se\ rechaza\ Ho$

Donde:

$$F_\alpha = 3.9207$$
$$F_c = 1.1201$$

Por lo tanto:

$F_c = 1.1201 < F_\alpha = 3.9207$ indicando que se acepta la Ho, y por tanto, se concluye, a partir de esta prueba, que el modelo presenta homoscedasticidad.

Continuando con la evaluación de los valores de F, se procede a contrastar la probabilidad asociada a F, bajo la siguiente regla de decisión:

$Si\ Prob.\ asociada\ a\ F\ es < 0.05 \qquad Se\ rechaza\ Ho$
$Si\ Prob.\ asociada\ a\ F\ es \geq 0.05 \qquad Se\ acepta\ Ho$

Como la Probabilidad asociada a F es igual a 0.2920 > 0.05, se acepta la hipótesis nula, es decir, el modelo presenta homoscedasticidad incondicional autorregresiva.

Por lo que concierne al R^2 éste es igual a 0.0093 = 0.93%, es decir, los residuales al cuadrado tienen una ínfima relación con ellos mismos rezagados 1 periodo, lo cual corrobora que el modelo no presenta heteroscedasticidad.

Por tanto, concluimos que el nuevo modelo planteado no presenta problemas de heteroscedasticidad de acuerdo con los resultados obtenidos a partir de cada una de las pruebas realizadas con anterioridad.

VI.6.- Reactivos para reafirmar los conocimientos.

A- EXAMEN AUTO EVALUATORIO: Afirmación de conceptos

Tema: Efecto sobre los estimadores de la violación de los supuestos del modelo de regresión lineal: Mínimos cuadrados ordinarios, HETEROSCEDASTICIDAD.

A.-Conteste con una “X” en **SI** cuando la afirmación sea verdadera y también con una “X” en NO cuando la afirmación sea falsa:

1.-Uno de los supuestos es que la varianza de las Ui no es constante: Si___; No____.

2.-Para verificar si hay o no heterocedasticidad se usan métodos gráficos y numéricos: SI_____;NO_____.

3.- La prueba White no forma parte de los métodos numéricos para identificar la heteroscedasticidad: Si____,NO____

4.- Con los métodos numéricos se establece la hipótesis nula de varianza variable y la hipótesis alternativa de varianza constante: Si_____; No________

5.-Se recomienda hacer prueba de Goldfeld y Quandt cuando las muestras son chicas : SI:_______; NO:___________

6.- Para aplicar la prueba de Goldfeld y Quandt los datos se deben de clasificar en dos grupos, en el primero se ponen los de valor pequeño y en el segundo los de valor mayor de Ui obtenidos con respecto a la variable regresada. **Enseguida se corre la regresión para cada uno de los dos grupos y, con los datos** *se hace la prueba F :* SI:__; NO__.

7.-La prueba F es un cociente que resulta de dividir la Varianza: suma de los residuos grandes al cuadrado por la varianza:suma de los residuos chicos al cuadrado : SI___; NO____.

8.- Los grados de libertad: en F, se obtienen con **n:** número de parámetros, **k** es el número de observaciones y **d** es un número escogido arbitrariamente: SI____; NO______:

9.- Cuando la F real es mayor que la F teórica se acepta la hipótesis alternativa de que hay heteroscedasticidad: SI____;NO_____.

10.- Para resolver el problema de heteroscedasticidad con la prueba de Goldfeld y Quandt los datos de las variables se transforman en una función Log Log :SI_____; NO____.

11.-Al eliminarse la heteroscedasticidad los estimadores de mínimos cuadrados que se obtienen son eficientes y consistentes, además de insesgados: SI____; NO______.

B- Ejercicio adicional*: Identificación de la heteroscedasticidad en los estimadores utilizando el Programa Eviews 5.*

Planteamiento:

Los siguientes son los gastos de consumo, Cy, e Ingreso disponible, Ix, de 30 familias, mensualmente.

Consumo mensual, Cy. en pesos	**Ingreso disponible mensual, Ix, en pesos**
10600	12000
11400	13000
12300	14000
13000	15000
13800	16000
14400	17000
15000	18000
15900	19000
16900	20000
17200	21000
10800	12000
11700	13000
12600	14000
13300	15000
14000	16000
14900	17000
15700	18000
16500	19000
17500	20000
17800	21000
11100	12000
12100	13000
13200	14000
13600	15000
14200	16000
15300	17000
16400	18000
16900	19000
18100	20000
18500	21000

1:- Con el método de White y Goldfeld & Quant *identifique* si hay heteroscedasticidad;

Con cualesquiera de los métodos aquí expuestos. Elimínela con cualquiera de los métodos conocidos.

***VI.2.- Autocorrelación*.**

La autocorrelación se puede presentar debido a factores como la inercia o la lentitud de las series económicas, al sesgo de especificación, a variables excluidas en la estimación, por utilizar la forma funcional incorrecta, debido a la manipulación de datos, es decir por el manejo y transformación de los datos (Gujarati, 2004: 471).

Así, se dice que la autocorrelación es típica de las series de tiempo debido a que el efecto del valor de un término, digamos de ayer, incide en el valor del término de hoy. Si hablamos en el lenguaje de la hipótesis nula, ésta se establece diciendo que los términos de error (u_i) en el modelo de regresión son independientes, es decir: **Ho: r=0**, lo cual indica que no hay correlación entre ellos.

Lo contrario, es hacer mención a la hipótesis alternativa indicando que es el relajamiento de este supuesto (hipótesis nula), es decir **Ha: r distinto de cero**, donde r es el coeficiente de correlación entre las U_i, lo cual indica que dichos términos de error, son dependientes unos de otro. Lo anterior significa que hay relación entre las U_i, que están correlacionadas, mismas que vistos en el ámbito de las SERIES DE TIEMPO, revelan que hay AUTOCORRELACION entre ellas, la cual se expresa en este caso con **ρ**, cuyo significado ahora es que hay autocorrelación entre el valor del término actual de la serie Ut y el anterior U_{t-1}. Tanto r como ρ toman valores que van de -1 a +1, con la interpretación ya conocida. Ejemplo, si analizamos el ingreso de las personas en varios años, el ingreso del año uno influye en el ingreso del año dos, este en el del año tres, y así sucesivamente, esto origina una autocorrelación de las Ui en el tiempo, que en series de tiempo se conoce como autocorrelación serial.

VI.2.1.- Identificación de autocorrelación: se hace con la (r) para datos de corte transversal y con (ρ) para series temporales, así como con la estadística d de Durbin-Watson y la h de durbin y otros metodos que se ilustran a continuación.

a) Aquí como en la heteroscedasticidad se usa r; de manera que en principio de manera general se dice que cuando su valor es alto: cercano a más uno o a menos uno, se dice que hay autocorrelación entre las Ui. Ahora usando la D, decimos:

b) Prueba d de Durbin y Watson.

Como el término de error (u_t) de un año está autocorrelacionado con el del año inmediato anterior (u_{t-1}), Durbin y Watson elaboraron la estadística "**d**", que sirve para detectar la autocorrelación y se determina con la fórmula:

$$d = \frac{\sum_{2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{1}^{n}\hat{u}_t^2}$$

en la que $\hat{u}_t$ se define como el residuo estimado para el período o año t.

Si desarrollamos el cuadrado de la fórmula de d, obtenemos

$$d = \frac{\sum\hat{u}_t^2 + \sum u_{t-1}^2 - 2\sum\hat{u}_t\hat{u}_{t-1}}{\sum\hat{u}_t^2}$$

Tomando en cuenta que cuando la muestra es grande se observa que $\sum \hat{u}_t^2$ y $\sum \hat{u}_{t-1}^2$ son casi iguales ya que difieren en una observación, tal que podemos decir 1+1= 2; en otras palabras ambas sumatorias son iguales, y si factorizamos tenemos que d= 2(1- la segunda parte del desarrollo), dividida entre el denominador que ahí aparece; luego, si decimos que r representa la autocorrelación entre ellos, es decir que r representa la segunda y última parte de la ecuación, entonces podemos establecer que la fórmula se puede expresar como:

d $\cong$ 2(1 - r), al respecto recuerde que r en series temporales se expresa como ρ.

Ahora bien, puesto que sabemos que r ó ρ oscilan entre –1 y +1, con desigualdades podemos decir para r lo siguiente: $-1 \leq r \leq +1$ y para ρ lo mismo.

Derivado de lo anterior, podemos establecer las siguientes igualdades:

Cuando r ó ρ = + 1, se dice que d = 0; hay autocorrelación positiva entre U_t y U_{t-1};

Cuando r ó ρ = -1, se dice que d = 4; hay autocorrelación negativa entre U_t y U_{t-1} ; y

Cuando r ó ρ = 0, se dice que d = 2; no hay autocorrelación entre U_t y U_{t-1} .

Por consiguiente cuando d tenga valores cercanos a 0 o 4, diremos que los residuos U_t y U_{t-1} están altamente correlacionados y que sus valores son dependientes entre si.

Es importante decir que la distribución muestral de d depende del valor de las variables explicativas. Durbin y Watson calcularon los valores de los limites superior (d_u) e inferior (d_L) para diferentes niveles de significación de d. Estos valores están en tablas mediante las cuales se prueban hipótesis nulas: autocorrelación cero versus las hipótesis alternativas: autocorrelación positiva de primer orden (entre u_t y u_{t-1}); cuando la autocorrelación es negativa se intercambian d_u y d_L. Luego si:

d < d_L, se rechaza la hipótesis nula de no autocorrelación, hay autocorrelación, debe corregirse.

d > d_u , no se rechaza la hipótesis nula de independencia, no se hace nada.

d_L < d < d_u , la prueba no es concluyente, es decir no sabemos si los términos de error u_i están autocorrelacionados o son independientes.

Lo anterior dicho en palabras de Dominick Salvatore(9): (“Econometría” Editorial Mc Graw Hill, página 153).

Si d < d_L, se acepta la hipótesis de autocorrelación, Ha: r≠ 0 **y se rechaza Ho: r=0**

d > d_u, se rechaza la hipótesis de autocorrelación, Ha: r≠ 0 **y se acepta Ho: r=0**

Para probar la Ho se compara la d calculada con la d en tablas partiendo de que está demostrado que la esperanza matemática de d, cuando r ó ρ = 0, está dada por la fórmula:

$$E(d) = 2 + \frac{2(k-1)}{n-k}$$

K es igual al número de parámetros de regresión estimados (se incluye el término constante). Dominick Salvatore[9] dice que k = número de variables explicativas + 1 (término constante), ver Anexo.5 en el anexo de todas las tablas estadísticas, y si n es el tamaño de la muestra, vamos a A.5 y encontramos $k^{'1}$, que necesitamos para obtener diferentes valores de d. Con estos datos se buscan en la tabla de Durbin Watson los valores d_L y d_u y se comparan con la d calculada para identificar si hay o no autocorrelación entre los residuos.

VI.2.2.- Orden de autocorrelación.

Por otra parte, también es muy importante indicar que una vez que sabemos que hay autocorrelación, el siguiente **paso es identificar de qué orden es la autocorrelación,** ello se sabe al calcular y trazar el **correlograma,** de manera que si vemos en él que la gráfica de la función de autocorrelación parcial, por ejemplo, sólo muestra **una** “barra” que se salga del intervalo de confianza, entonces decimos que la autocorrelación es de primer orden , y se expresa así: AR(1), (Pérez, 2007: 176)

VI.2.3.- Consecuencia de la autocorrelación.

Como indica Dominick Salvatore[9] , la presencia de autocorrelación es común en “Series de Tiempo y lleva a errores estándar sesgados hacia abajo (y así a pruebas estadísticas e intervalos de confianza incorrectos)”. Gujarati (1991: 298) por su parte dice que “aun cuando los estimadores MCO continúan siendo lineales, insesgados y consistentes, pero dejan de ser eficientes”, situación que provoca las mismas consecuencias que Salvatore señaló.

VI.2.4.- Corrección de autocorrelación.

Por lo extenso de la metodología que existe para corregir la autocorrelación, decidí presentarla en los siguientes ejercicios donde hay un planteamiento integral, es decir, además de plantear el origen de la autocorrelación, también se describe e ilustra la metodología para resolverla. Así, aquí en esta sección sencillamente me concreto a enunciar su solución comentando que:

a) Dominick Salvatore (*) dice que para corregir la autocorrelación **se debe estimar ρ, por ser el indicador de la autocorrelación serial.** Así, se determina a partir de d= 2(1-r)*;* despejando obtenemos r=2-d/2, de manera que cuando por ejemplo d=0.88, vemos que r= 2-0.88/2=1.12/2=0.56, valor a utilizar para reducir o eliminar la autocorelación. Así, según el valor que tome r será la reducción o eliminación de la autocorrelación (Gujarati,1991:323).

b) El mismo autor Gujarati (1991: 330) comenta que **Theil y Nagar** sugieren que en *muestras pequeñas* r (nuestro ρ) se debe estimar con la fórmula:

$$r = \frac{N^2(1-d/2)+k^2}{N^2-k^2}$$

Ejemplo numérico trabajado manualmente para corregir la autocorrelación

D. Salvatore presenta el nivel de inventarios, Y, y ventas X, los dos en miles de millones de dólares para la industria de manufacturas de los E.E. U.U., del año 1959 al de 1978. Hace la regresión de Y con X con los siguientes datos:

Año	**1959**	**1960**	**1961**	**1962**	**1963**	**1964**	**1965**	**1966**	**1967**	**1968**	**1969**	**1970**	**1971**	**1972**	**1973**	**1974**	**1975**	**1976**	**1977**	**1978**
Y	52.9	53.8	54.9	58.2	60.0	63.4	68.2	78.0	84.7	90.6	98.2	101.7	102.7	108.3	124.7	157.9	158.2	170.2	180.0	198.0
X	30.3	30.9	30.9	33.4	35.1	37.3	41.0	44.9	46.5	50.3	53.5	52.8	55.9	63.0	73.0	84.8	86.6	98.8	110.8	124.7

Obtiene $y_t = 6.61 + 1.63 x_t$ $R^2 = 0.98$

(1.98) (32.0) d = 0.69

(3.33) (0.05) hecho con Eviews

Que en detalle es:

Dependent Variable: Y
Method: Least Squares
Date: 11/22/04 Time: 21:13
Sample(adjusted): 1959 1978
Included observations: 20 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	**6.608085**	**3.329150**	1.984917	0.0626
X	**1.631438**	**0.050975**	32.00487	0.0000

R-squared	**0.982731**	Mean dependent var	103.2300
Adjusted R-squared	0.981771	S.D. dependent var	46.47964
S.E. of regression	6.275390	Akaike info criterion	6.605788
Sum squared resid	708.8494	Schwarz criterion	6.705361
Log likelihood	-64.05788	F-statistic	1024.312
Durbin-Watson stat	**0.696772**	Prob(F-statistic)	0.000000

Dado que con n = 20 y K[1] =2-1= 1 y ~ = 5% $d_L = 1.20$ tenemos que d = 0.70 < $d_L = 1.20$ se acepta la hipótesis de autocorrelación. Así, para *corregir la autocorrelación, se dice que una estimación de r* esta dada por r= 2-d/2 = 2-0.70/2= 1.30/2=0.65
Con la otra fórmula, de Gujarati, se obtiene r= 0.67

Si usamos r=0.67 para transformar las variables originales y utilizando el dato del año de 1959 : 52.9 $\sqrt{1 - (0.67)^2}$ = 39.27 y del mismo año el valor de las ventas, 30.3 $\sqrt{1-(0.67)^2} = 22.49$ para la primera observación **transformado de Y y X**, respectivamente. Para el resto de los valores transformados de Y e X se calcula de la siguiente manera:

Puesto que con r= 0.67 obtenemos r cuadrada= 0.4489, entonces usamos $y_1\sqrt{1 \text{ J } \hat{r}^2}$ para el primer dato de Y, que corresponde a 1959, y para no desecharlo

$Y_1^* = Y_1\sqrt{1-r^2} = 52.9\sqrt{.5511} = 52.9(.74) = 39.27$ para el primer término de Y

$Y_2^* = Y_2 - rY_1 = 53.8 - 0.67(52.9) = 53.8 - 35.44 = 18.36$ para el segundo y subsecuentes Y´s, ver ecuaciones

$Y_3^* = Y_3 - rY_2 = 54.9 - 0.67(53.8) = 18.85$

$Y_4^* = Y_4 - rY_3 = 58.2 - 0.67(54.9) = 21.41$

$Y_5^* = Y_5 - rY_4 = 60.0 - 0.67(58.2) = 21.01$

$Y_6^* = Y_6 - rY_5 = 63.4 - 0.67(60.0) = 23.20$

$Y_7^* = Y_7 - rY_6 = 68.2 - 0.67(63.4) = 25.72$

$Y_8^* = Y_8 - rY_7 = 78.0 - 0.67(68.2) = 32.31$

$Y_9^* = Y_9 - rY_8 = 84.7 - 0.67(78.0) = 32.44$

$Y_{10}^* = Y_{10} - rY_9 = 90.6 - 0.67(84.7) = 33.85$

$Y_{11}^* = Y_{11} - rY_{10} = 98.2 - 0.67(90.6) = 37.50$

$Y_{12}^* = Y_{12} - rY_{11} = 101.7 - 0.67(98.2) = 35.91$

$Y_{13}^* = Y_{13} - rY_{12} = 102.7 - 0.67(101.7) = 34.56$

$Y_{14}^* = Y_{14} - rY_{13} = 108.3 - 0.67(102.7) = 39.49$

$Y_{15}^* = Y_{15} - rY_{14} = 124.7 - 0.67(108.3) = 52.14$

$Y_{16}^* = Y_{16} - rY_{15} = 157.9 - 0.67(124.7) = 74.35$

$Y_{17}^* = Y_{17} - rY_{16} = 158.2 - 0.67(157.9) = 52.41$

$Y_{18}^* = Y_{18} - rY_{17} = 170.2 - 0.67(158.2) = 64.21$

$Y_{19}^* = Y_{19} - rY_{18} = 180.0 - 0.67(170.2) = 65.97$

$Y_{20}^* = Y_{20} - rY_{19} = 198.0 - 0.67(180.0) = 77.40$

Hacemos lo mismo para la transformación de las X´s con r=0.67 y r cuadrada=0.4489 usamos $x_1\sqrt{1 \text{J} \hat{r}^2}$ para el primer dato de X, que corresponde a 1959, y para no desecharlo

$X_1^* = X_1\sqrt{1-r^2} = 30.3\sqrt{1-0.4489} = 30.3\sqrt{0.5511} = 30.3(0.74) = 22.49$ para el primer término de X

$X_2^* = X_2 - rX_1 = 30.9 - 0.67(30.3) = 10.60$;para el segundo y subsecuentes X´s, seguir ecuaciones

$X_3^* = X_3 - rX_2 = 30.9 - 0.67(30.9) = 10.20$

$X_4^* = X_4 - rX_3 = 33.4 - 0.67(30.9) = 12.70$

$X_5^* = X_5 - rX_4 = 35.1 - 0.67(33.4) = 12.72$

$X_6^* = X_6 - rX_5 = 37.3 - 0.67(35.1) = 13.78$

$X_7^* = X_7 - rX_6 = 41.0 - 0.67(37.3) = 16.00$

$X_8^* = X_8 - rX_7 = 44.9 - 0.67(41.0) = 17.43$

$X_9^* = X_9 - rX_8 = 46.5 - 0.67(44.9) = 16.42$

$X_{10}^* = X_{10} - rX_9 = 50.3 - 0.67(46.5) = 19.15$

$X_{11}^* = X_{11} - rX_{10} = 53.5 - 0.67(50.3) = 19.80$

$X_{12}^* = X_{12} - rX_{11} = 52.8 - 0.67(53.5) = 16.96$

$X_{13}^* = X_{13} - rX_{12} = 55.9 - 0.67(52.8) = 20.52$

$X_{14}^* = X_{14} - rX_{13} = 63.0 - 0.67(55.9) = 25.55$

$X_{15}^* = X_{15} - rX_{14} = 73.0 - 0.67(63.0) = 30.79$

$X_{16}^* = X_{16} - rX_{15} = 84.0 - 0.67(73.0) = 35.89$

$X_{17}^* = X_{17} - rX_{16} = 86.6 - 0.67(84.8) = 29.78$

$X_{18}^* = X_{18} - rX_{17} = 98.8 - 0.67(86.6) = 40.78$

$X_{19}^* = X_{19} - rX_{18} = 110.8 - 0.67(98.8) = 44.60$

$X_{20}^* = X_{20} - rX_{19} = 124.7 - 0.67(110.8) = 50.46$

Con los datos nuevos, transformados de Y e X, a partir de r, ahora corremos nuevamente la regresión sobre las variables transformadas (*que identificaremos con* *), sin omitir los datos de 1959, y se obtienen:

$$y_t^* = 4.65 + 1.52\,x_t^* \qquad R^2 = 0.94$$

$$(2.42) \quad (0.08) \qquad d = 1.32$$

De manera detallada usando Eviews:

Dependent Variable: YCALC
Method: Least Squares
Date: 11/20/04 Time: 14:32
Sample: 1959 1978
Included observations: 20

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.656444	2.423902	1.921053	0.0707
XCALC	1.526495	0.089229	17.10764	0.0000

R-squared	0.942061	Mean dependent var	41.93650
Adjusted R-squared	0.938842	S.D. dependent var	19.19455
S.E. of regression	4.746834	Akaike info criterion	6.047472
Sum squared resid	405.5838	Schwarz criterion	6.147046
Log likelihood	-58.47472	F-statistic	292.6714
Durbin-Watson stat	1.327927	Prob(F-statistic)	0.000000

Vemos en la tabla de Durbin y Watson que con ~ = 5%, n = 20 y K^1 = 1 se obtiene d_U = 1.41 y d_L=1.20. <u>Por consiguiente decimos que d = 1.32, está entre estos dos valores anteriores, lo cual significa que la autocorrelación esta indefinida.</u>
Por otra parte, es interesante señalar que cuando **omitimos** los datos de Y e X <u>del primer año, **1959,** al correr la ecuación de regresión se obtiene el siguiente valor de d cuyas “estadísticas” no difieren sustancialmente de la anterior, como se muestra en los siguientes resultados:</u>

Dependent Variable: YTRNSF
Method: Least Squares
Date: 11/21/04 Time: 09:21
Sample(adjusted): 1960 1978
Included observations: 19 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.526473	2.479092	1.825859	0.0855
XTRNSF	1.519796	0.094774	16.03594	0.0000

R-squared	0.937990	Mean dependent var	40.05211
Adjusted R-squared	0.934343	S.D. dependent var	18.92771
S.E. of regression	4.849969	Akaike info criterion	6.095122
Sum squared resid	399.8774	Schwarz criterion	6.194537
Log likelihood	-55.90366	F-statistic	257.1515
Durbin-Watson stat	1.335496	Prob(F-statistic)	0.000000

VI.2.4.1.- Identificación de la presencia o ausencia de autocorrelación con Variables Retardadas o Rezagadas.

Cabe señalar que su aplicación es fundamentalmente en series de tiempo en virtud de que su uso se sustenta en la observación de que el impacto de algunas variables exógenas sobre las endógenas no se concreta u observa solamente en un punto en el tiempo (manejo de datos de corte transversal), <u>sino que puede prolongarse a un lapso mucho mayor</u>; por ejemplo: el ingreso de las personas del año pasado influye en su ingreso de este año; también, la inflación del mes pasado incide en la inflación de este mes, etc.

En este contexto es que decimos que los efectos diferidos en el tiempo se incorporan en un modelo econométrico mediante el uso de “variables retardadas o rezagadas” . Dichos efectos pueden indicar autocorrelación o dependencia entre los términos de error U_i de los años sucesivos de la serie bajo estudio. Al respecto, dichos efectos pueden expresarse de dos maneras diferentes, las cuales Son:

VI.2.4.1.1.- Por medio de modelos autorregresivos.

En este caso la incidencia del tiempo (t) se cuantifica por medio del uso de la variable endógena Y_t ahora como variable exógena pero retardada en algún lapso de tiempo, digamos anualmente, misma que se denota como Y_{t-1} o Y(-1), donde (-1) se conoce como rezago o retardo de un año; igualmente, en ese sentido digamos Y_{t-3} o Y(-3) indica que Y ha sido rezagada o retardada tres años, la cual ahora puede usarse como variable explicativa de Y en algún periodo de análisis determinado dado que las Y_{t-1} Y_{t-3} son las variables que ahora mantienen sus efectos en Y.

Por medio de modelos de rezagos o retardos distribuidos.

A diferencia de los modelos autorregresivos, en estos las variables explicativas X_1 y X_2 son las que mantienen impactos o efectos prolongados en el tiempo sobre la variable dependiente Y.

VI.2.4.1.2.- Modelos autorregresivos usados para detectar autocorrelación con h llamada “ Contraste h de Durbin”, definiendo a Y como consumo, X_1: ingreso, X_2: inflación, de los últimos 30 meses, cuya base de datos es la siguiente:

Observación	Y	X_1	X_2
1	3	1	8
2	2	2	15
3	4	2.5	10
4	5	3	9
5	5	4	7
6	7	5	6
7	6	7	8
8	8	8	4
9	9	9	3
10	12	15	1
11	12.2	16	1.5
12	12.4	13.5	4
13	12.1	13.6	3.5
14	12.5	14.2	4.6
15	12.9	16.5	2.3
16	13.8	16.2	4.2
17	14.1	18.6	3.9
18	13.1	19.3	5.2
19	13	17.2	4.1
20	13.2	18.6	3.19
21	13.3	15.6	4.6

22	13.4	19.2	3.9
23	13.5	13.6	4.2
24	14.1	15.6	4.5
25	14.2	17.5	3.1
26	14.3	16.2	5.3
27	14.4	14.2	4.2
28	15	16.2	4.3
29	15.2	16.6	3.8
30	15.6	17.5	4

En opinión de Carrascal et al (2001: 294) si usamos los modelos autorregresivos es porque "suponen una violación de una de las hipótesis establecidas en el modelo de regresión lineal clásico. Concretamente, *la hipótesis de que los regresores no son aleatorios*, ya que la variable endógena retardada depende de la perturbación aleatoria y, por tanto, tiene un carácter estocástico". *Es por eso que cuando hay regresores estocásticos: Y (-1)* que la estadística **d** de Durbin Watson ya no es apropiada para identificar la autocorrelación que pueda existir entre la variable endógena Y (-1) retardada y la perturbación aleatoria (Ui).

César Pérez (2007: 129) es más parco, no dice porqué usar **h,** simplemente informa que " El estadístico de Durbin Watson no debe utilizarse para modelos que introducen retardos en la variable dependiente ni para modelos sin término constante."

Por consiguiente, continuando con Carrascal et al, es importante mencionar que con este método las estimaciones están en función del tipo de **dependencia en el tiempo** que se detecte entre la variable estocástica Y(-1) y las Ui´s, es decir, depende del tiempo que dure su efecto sobre Y. Cuando es parcial (en tiempo pasado pero no en el presente y menos en el futuro) se dice que los estimadores mantienen su propiedad de consistencia pero pierden la de insesgabilidad, pero aun así pueden seguirse usando. Sin embargo, cuando la dependencia es total (siempre) pierden las dos propiedades antes mencionadas y se busca un **método** de estimación alterno que puede ser el de **variables instrumentales,** ya que con éste se garantiza la propiedad asintótica de consistencia en las estimaciones con mínimos cuadrados ordinarios (MCO). En consecuencia, se infiere que cuando no hay dependencia (incorrelación) de Y(-1) de las Ui´s los estimadores conservan las dos propiedades. En este sentido se indica que cuando no hay correlación entre ellas se dice que existe *dependencia parcial* , en tanto que cuando se detecta que hay, se dice que existe una *dependencia total.*

C.1:1.- Contraste **h** de Durbin

Como se hizo con la **d** de Durbin Watson, el contraste **h** de Durbin establece las mismas hipótesis:

Ho: $\rho=0$, no hay correlación de Y(-1) con las Ui´s

Ha: $\rho\geq0$ ó $\rho\leq0$,hay correlación AR(1) positiva o negativa entre ellas.

La fórmula para calcular manualmente h= ° *la raíz cuadrada del cociente de T/1-

$TS^2_{\beta k+1}$. Se establece que h se distribuye normalmente con media aritmética cero y varianza 1; donde ρ es el estimador obtenido en la regresión que se corre de las Ui frente así mismas retardadas un periodo y sin término constante; $TS^2_{\beta k+1}$ es el valor de la varianza del estimador correspondiente a la variable endógena retardada un periodo Y(-1) y finalmente T: es el número de términos (Carrascal et al, 2001: 296). Aquí es importante observar que el **signo** de la estadística **h** en la Ha es igual al de ρ. Ejemplo: con α=5% trabajada en una prueba de una cola o extremo, cuando h es mayor que $t_{\alpha} = +1.645$ se rechaza Ho de incorrelación y se acepta Ha de que existe autocorrelación positiva entre Y(-1) y las Ui´s; luego cuando **h** es menor que $t_{\alpha} = -1.645$ se rechaza Ho frente a la aceptación de Ha de que se ha detectado autocorrelación negariva.

Así, para detectar con **Eviews** si hay autocorrelación los pasos a seguir son los siguientes:

1. Crear la ecuación: Quick/Estimate Ecuation/ Y_C_X1_Y(-1) / ok. Cuyos resultados son los siguientes:

Dependent Variable: Y
Method: Least Squares
Date: 10/28/08 Time: 12:25
Sample(adjusted): 2 30
Included observations: 29 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.424546	0.460451	3.093805	0.0047
X1	0.161815	0.088443	1.829606	0.0788
Y(-1)	0.715366	0.119971	5.962817	0.0000

R-squared	0.956057	Mean dependent var	11.38966
Adjusted R-squared	0.952677	S.D. dependent var	3.818035
S.E. of regression	0.830570	Akaike info criterion	2.564289
Sum squared resid	17.93603	Schwarz criterion	2.705734
Log likelihood	-34.18219	F-statistic	282.8386
Durbin-Watson stat	2.148425	Prob(F-statistic)	0.000000

Observe que estos resultados son para 29 datos porque hay un retardo o rezago, por eso leemos en "Sample (adjusted): 2 30" y en "included observatios: 29 after adjusting endpoints".

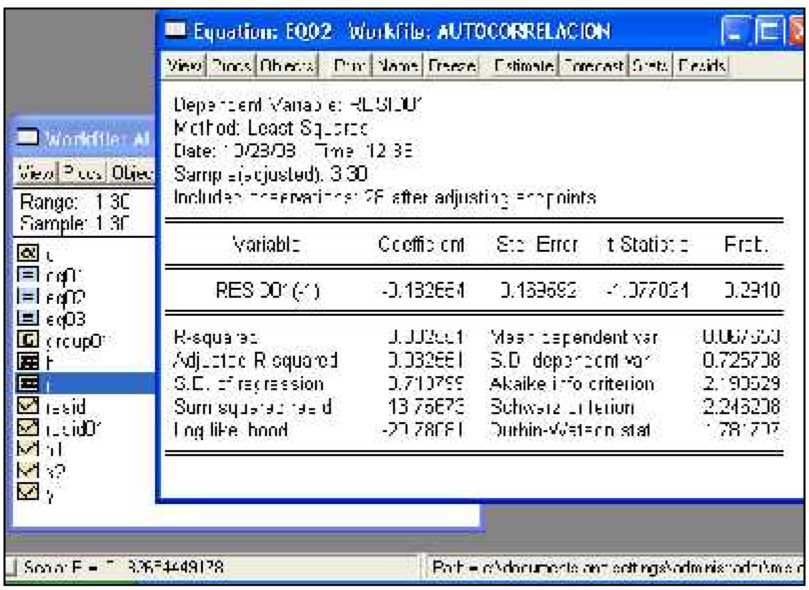
2. Crear la variable de los residuos (Ui) mediante: Procs/make residual series/ y en la ventana que se abre aceptamos "Residual Type": o "Ordinary" / ok y aparecen las Ui con el nombre de "Resid01".

3. Realizar la ecuación de las Ui creadas con respecto a ellas mismas rezagadas un periodo: Quick/Estimate Ecuation/ resid01 resid01(-1)/ ok y aparece el siguiente Cuadro:

Dependent Variable: RESID01
Method: Least Squares
Date: 10/28/08 Time: 12:33
Sample(adjusted): 3 30
Included observations: 28 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
RESID01(-1)	**-0.182654**	0.169592	-1.077024	0.2910
R-squared	0.032551	Mean dependent var		0.067653
Adjusted R-squared	0.032551	S.D. dependent var		0.725708
S.E. of regression	0.713799	Akaike info criterion		2.198629
Sum squared resid	13.75673	Schwarz criterion		2.246208
Log likelihood	-29.78081	Durbin-Watson stat		1.781797

En este caso el dato importante, y por tanto el que vamos a rescatar, es el valor del coeficiente de la variable retardada “Resid01(-1)” que identificamos con la letra “ρ” (rho) y que en Eviews representamos con “r” que es igual a: -0.182654, mismo que en EViews se guarda como objeto escalar de la siguiente forma: vamos al menú de comandos de la ventana principal y ahí escribimos: scalar_r=@coefs(1)/enter y primero aparece como un objeto más en el Workfile mismo que si le damos doble click su valor aparecerá en la parte inferior izquierda de la pantalla como se muestra a continuación:



4. Se calcula la “h de Durbin” escribiendo en la ventana de comandos lo siguiente:

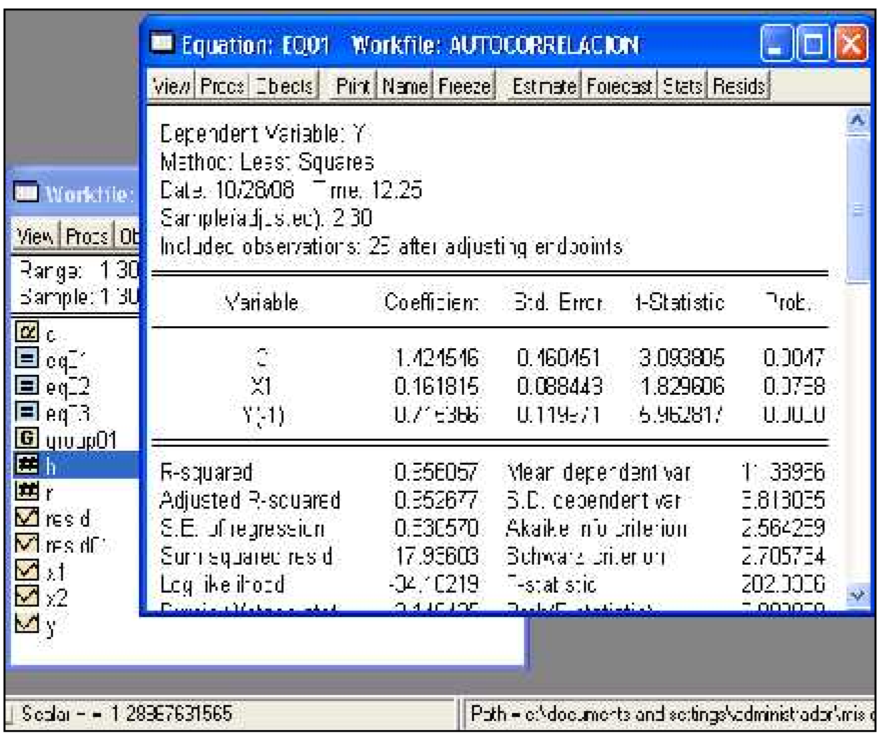
Scalar_h=r*@sqrt(@regobs/(1-@regobs*(@stderrs(3))^2))

Nota: Para calcular “h” necesitamos que esté activada la ecuación original, aquella en donde está incluida la variable endógena retardada como exógena, esto se logra dando un click sobre ella.

Donde:

r = Coeficiente de las Ui rezagadas un periodo.
Regobs = Número de observaciones, en este caso 30.
Stderrs(3) = Es la desviación estándar del estimador correspondiente a la variable endógena retardada Y(-1). El número 3 del stderrs corresponde a la posición de este coeficiente en la ventana de la ecuación de regresión en Eviews, por lo que puede variar según su posición o lugar que ocupe en dicha ventana: en nuestro caso ello sucediò cuando creamos en el paso 1 la ecuación de regresión en que establecimos Y(-1) en tercer lugar, por ello su stderrs ahora lleva el número 3. ; al respecto, dicha posición o lugar la establece subjetivamente la persona que está calculando h . En este contexto el 2 que le sigue en la ecuaciòn anterior de arriba para h, significa que stderrs fue elevado al cuadrado, i.e., indica que es la varianza de Y(-1). Cabe señalar que para elevar al cuadrado en Eviews hacemos: apretamos la tecla Alt simultáneamente con las teclas 9 y 4, enseguida se sueltan las tres y aperece en pantalla “^2”.

Así, en el ángulo inferior izquierdo aparece la leyenda “h successfully created”, para poder ver su valor damos doble click en el objeto *escalar* del Workfile que denominamos “h”.



Interpretación de “h de Durbin”: Su valor de -1.28867631565 se compara con los valores que toma Zα en una distribución normal que tenga media cero y varianza unitaria**,** cuando α= 5% en una prueba de una cola o extremo, tal que en ese punto Z_{α} = ±1.645, valores que contrastamos con h = - 1.288676, de manera que rechazamos Ho de ausencia de autocorrelación si el valor de h es mayor que los valores de la distribución en Z_{α} = ±1.645 al 95% de confianza. Como h está en la zona de aceptación, aceptamos Ho y decimos que no hay autocorrelación entre Y(-1) y los residuos Ui.

Atención: Si hubiéramos rechazado Ho, es decir aceptado Ha, estaríamos frente a un problema de autocorrelación, mismo que deberíamos de **resolver** llevando a cabo la estimación incorporando un esquema AR(1) y partiendo de la estimación por **variables instrumentales** para ver si es total la dependencia de Y(-1) de las Ui´s. Con esta metodología obtendremos estimadores consistentes de todos los parámetros incluido el del esquema AR(1). Así, **suponiendo** que hubiera autocorrelación entre Y(-1) y Ui´s, ilustremos cómo se eliminaría con el siguiente ejemplo:

VI.2.4.1.3.- Método con variables instrumentales.

En este caso debe cuidarse que el número de variables instrumentales debe ser menor o igual que el número de regresores en el modelo econométrico (el programa Eviews incluye por default una variables con valor de 1 cuando no se haya incluido C entre las mismas), por eso utilizamos X_1 y $X_1(-1)$.
1. Creamos la ecuación con los instrumentos X1 y X1(-1) de la siguiente manera: Quick/Estimate Equation/ y en la especificación del modelo escribimos: Y_X1_Y(-1)_C /, abajo en el método seleccionamos: TSLS-Two stage Least Squares (TSNLS and ARMA), luego en “Instrument List” escribimos: X1_X1(-1)/ok y aparece la siguiente ecuación:

Dependent Variable: Y
Method: Two-Stage Least Squares
Date: 10/28/08 Time: 13:14
Sample(adjusted): 2 30
Included observations: 29 after adjusting endpoints
Instrument list: X1 X1(-1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.209346	0.136012	1.539176	0.1358
Y(-1)	0.647204	0.190603	3.395554	0.0022
C	1.546159	0.532832	2.901779	0.0075

R-squared	0.955512	Mean dependent var	11.38966
Adjusted R-squared	0.952089	S.D. dependent var	3.818035
S.E. of regression	0.835710	Sum squared resid	18.15871
F-statistic	267.5755	Durbin-Watson stat	2.044055
Prob(F-statistic)	0.000000		

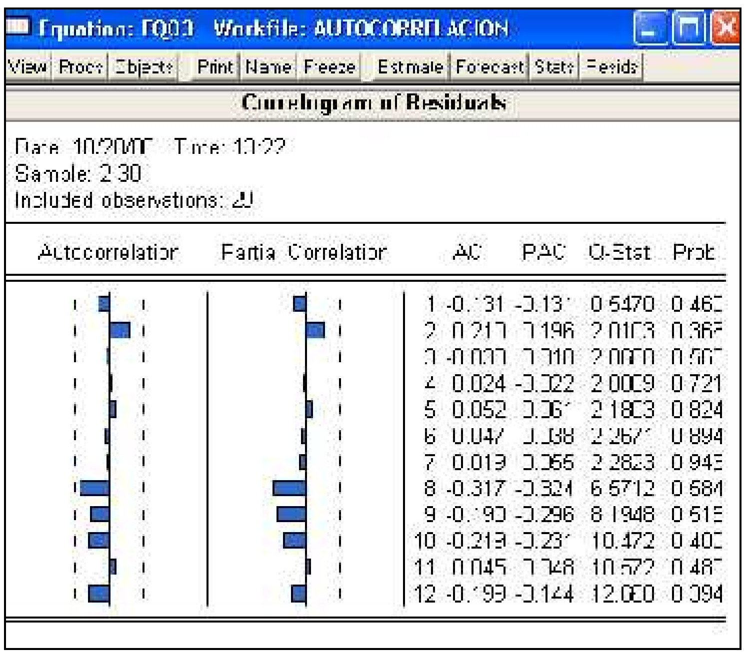
El método arriba sombreado se traduce al español como “mínimos cuadrados en dos etapas” , por ejecutar el caso especial de lo que estamos llamando “ variables instrumentales”; al hacerlo, obtiene los instrumentos porque obtiene en una primera etapa las regresiones por MCO de cada uno de los regresores de la ecuación lineal: X_1 e $X_1(-1)$ y luego, en una segunda etapa, señala Carrascal et al (2001:300) “efectúa la estimación por mínimos cuadrados ordinarios de la variable endógena original (Y) utilizando como variables explicativas los valores estimados de cada una de las regresiones de la primera etapa. Se demuestra que estas dos etapas son equivalentes a la obtención de los estimadores por variables instrumentales cuando se utilizan como instrumentos los valores estimados de las regresiones de la primera etapa”.
Al respecto, en el Cuadro anterior se muestra sombreados el método y las variables instrumentales utilizadas para hacer la regresión lineal. Aquí conviene precisar y aclarar que las estadísticas se calcularon con las Ui´s provenientes de

las variables instrumentales y de las variables originales del modelo econométrico y, por consiguiente, no con las variables obtenidas en la segunda etapa del algoritmo de cálculo.

Interpretación de la eliminación de la autocorrelación con variables instrumentales: Los coeficientes de X_1 y de Y(-1) indican que el Ingreso(X_1) tiene un efecto a corto plazo sobre el consumo de 0.209346 y a largo plazo de 0.59339= 0.209346/1-0.647204=0.59339, en otras palabras, si el ingreso aumenta en una unidad en el periodo, el consumo crece en 0.209346 en ese lapso. Esa unidad adicional en el ingreso produce efectos sobre el consumo de los siguientes periodos, tal que en el largo plazo su incidencia será de 0.59339 unidades. ¿Como acotar el largo plazo, es decir, qué número de años comprende?
La respuesta está en el estudio de los modelos autorregresivos vistos como modelos con retardos infinitos de la variable explicativa.

Comentarios: al haberse logrado la incorrelación, ésta se puede verificar con:

a).- Utilizando el correlograma, la estadística Q y la prueba Breusch – Godfrey para corroborar lo anterior. Para la primera: View/Residual Test/ Correlogram Q-satats y aparece:



	AC	PAC	Q-Stat	Prob
1	-0.131	-0.131	0.5470	0.46[illegible]
2	0.217	0.196	2.01[illegible]3	0.36[illegible]
3	-0.007	0.010	2.0[illegible]	0.5[illegible]
4	0.024	-0.022	2.00[illegible]9	0.721
5	0.052	0.06[illegible]	2.18[illegible]3	0.824
6	0.047	0.038	2.267[illegible]	0.894
7	0.019	0.055	2.2823	0.943
8	-0.317	-0.324	6.5712	0.584
9	-0.[illegible]93	-0.296	8.1948	0.515
10	-0.213	-0.23[illegible]	10.472	0.40[illegible]
11	0.045	0.048	10.572	0.48[illegible]
12	-0.[illegible]93	-0.144	12.0[illegible]0	0.394

Interpretación: Grafica y numéricamente vemos que los coeficientes de correlación no se salen de las bandas de confianza y que en AC como en PAC los valores de ° son bajos, que indica que no existe un problema serio de autocorrelación; igualmente vemos que la probabilidad de Q en todos los casos es mayor que 0.05 o 5% lo cual indica que no hay autocorrelación serial entre Y(-1) y las Ui.

b).- Con la prueba de Breusch-Godfrey: View/Residual Tests/Serial Correlation LM Test/ok y aparece:

Breusch-Godfrey Serial Correlation LM Test:				
Obs*R-squared	**0.611607**	**Probability**		**0.434183**
Test Equation:				
Dependent Variable: RESID				
Method: Two-Stage Least Squares				
Date: 10/28/08 Time: 13:27				
Presample missing value lagged residuals set to zero.				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
X1	0.046525	0.151170	0.307767	0.7608
Y(-1)	-0.062056	0.210086	-0.295386	0.7701
C	0.064875	0.544842	0.119071	0.9062
RESID(-1)	**-0.161584**	**0.220172**	**-0.733898**	**0.4698**
R-squared	**0.021090**	Mean dependent var		-2.68E-17
Adjusted R-squared	-0.096379	S.D. dependent var		0.805311
S.E. of regression	0.843226	Akaike info criterion		2.624278
Sum squared resid	17.77574	Schwarz criterion		2.812871
Log likelihood	-34.05203	F-statistic		0.179536
Durbin-Watson stat	1.709575	Prob(F-statistic)		0.909269

Interpretación: Observamos que p- valores: probabilidad para la Chi cuadrada "Obs*R-squared" es 0.434183 que es mayor que α = 5% e indica que no hay autocorrelación . Iguamente, la conclusión anterior también se alcanza al analizar la p-probabilidad del Coeficiente AR(1): **RESID(-1)** estimado que *no es significativo* porque su p: probabilidad de 0.4698 ≥ α=5%, lo que confirma la estructura AR(1) para los residuos, de manera que podemos decir que los errores muestran un esquema autorregresivo de orden 1 dada que la *no significación* estadística de RESID(-1) es 0.4698. Similarmente, dado que R^2 =0.021090 es un valor muy bajo, cercano a cero, ello también indica que hay incorrelación.

VI.2.4.1.3.1.-Sugerencias complementarias para la corrección de la autocorrelación con el manejo de variables retardadas en modelos autorregresivos.

Se dice que cuando se identifica que hay un problema de autocorrelación en los estimadores obtenidos con variables instrumentales, **ya que efectivamente ahora son consistentes pero ya no son eficientes**. Ello debe corregirse incorporando términos o esquemas AR (Auto regresiones) y MA (medias móviles) en el modelo por el método de variables instrumentales (Carrascal et al, 2001: 303)
Con Eviews el procedimiento es:
Quick/ estímate equation/ Y C Y(-1) AR(1) /Method:TSLN/ en variables instrumentales escribimos:X_1 X_1(-1) (-2)/ ok , enseguida debe de aparecer el Cuadro en que se muestre que ya no hay autocorrelación ,es decir, que la probabilidad del coeficiente de Y(-1) sea mayor al 5%. En congruencia con ello, también se pueden obtener el correlograma, la estadística Q y hacer la Prueba Breusch-Golfrey, cuyos resultados deben indicar que no hay autocorrelación.

Cuando esto no sucede, es decir, si no se hubiera corregido la autocorrelación, podríamos seguir usando AR: $X_1(-1)$, $X_2(-2)$,$X_3(-3)$,…..,X_M **tantas veces como fuera necesario para corregirla.**

***VI.2.4.1.3.1.1.- Modelos con retardos o rezagos distribuidos: finitos e infinitos*.**

Modelos con retardos finitos

El problema en determinar el número de retardos (¿ cuántos y porqué?) sabiendo que X_1 ejerce un efecto prolongado (la pregunta es: ¿ saber cuántos retardos es necesario hacer para conocer el efecto sobre Y?), se soluciona corriendo el modelo con X_1 retardada un *número M finito* de veces: ejemplo los M retardos se escriben en la caja de diálogo así: $X_1(-1)$, $X_2(-2)$,$X_3(-3)$,…..,$X_M(-M)$; por eso se le llama también "longitud del retardo". Luego que se obtienen los resultados de los diferentes modelos, se deben comparar y enseguida seleccionar aquel cuya R^2 ajustado sea mayor y que muestre el menor valor en los criterios de información de AKAIKE y SCHWARZ.

Con Eviews supóngase que si deseamos conocer el efecto prolongado de X_1: INGRESO sobre Y: CONSUMO. Así, si tenemos 12 años de datos para cada una de esas variables, empezamos a correr el modelo y hallamos la siguiente tabla:

M	T: Número de datos	R^2 ajustado	Criterio de AKAIKE	Criterio de SCHWARZ
0	12	0.7391	4.26	4.34
1	11	0.7766	4.03	4.14
2	10	0.8834	3.30	3.40
3	9	0.8937	3.15	3.22
4	**8**	**0.9151**	**2.91**	**2.94**
5	7	0.8436	3.44	3.42
6	6	0.7469	3.75	3.65

Interpretación: Supuestamente **el efecto prolongado** (Carrascal et al, 2001: 310 y 311) de X_1 sobre Y es de cuatro años.

VI.2.5.- APLICACIÓN CON EVIEWS.

Ejemplos utilizando EViews 5.

Antes de iniciar el tema, quiero aprovechar para informar que los métodos aquí utilizados complementan congruente y eficazmente los expuestos en la sección teórica anterior. Al término de esta sección contará con un acervo suficiente y actualizado para hacer análisis de autocorrelación.

Así, ahora diremos que la ilustración de la detección y corrección de la Autocorrelación se realizará mediante el uso de EViews utilizando datos actuales de la economía mexicana.

Ejemplo 1: Teoría del Consumo.

De acuerdo con la teoría Keynesiana el Consumo está en función del Ingreso. Por lo tanto, el Ingreso es igual al Consumo y la Inversión, con esto se tratará de hacer una aproximación para determinar cómo está explicado el Consumo Privado.

$C = f(Y)$

Puesto que el ingreso es igual a la suma del consumo más la inversión:

$Y = C + I$

Así, tomaremos los datos de la economía mexicana en el periodo de 1980 a 2010.

Entonces, para el caso de la economía mexicana, tomaremos el ingreso como el Producto Interno Bruto (PIB), a la Inversión como la Formación Bruta de Capital Fijo (FBKF) y se incluyó la variable Exportaciones, ello debido a que desde 1980 las exportaciones mexicanas han tenido gran impulso.

$CP = F(PIB + FBKF + X)$).

En términos de la regresión la presencia de la autocorrelación puede verse de la siguiente forma:

$E(ui, uj) \neq 0 \qquad i \neq j$

La prueba de hipótesis será la siguiente:

Ho: No autocorrelación.

Ha: Autocorrelación.

1. Causas del problema de Autocorrelación.

La autocorrelación se puede presentar debido a factores como: la inercia o la lentitud de las series económicas, al sesgo de especificación, variables excluidas en la estimación, por utilizar la forma funcional incorrecta, debido a la manipulación de datos, es decir por el manejo y transformación de los datos (Gujarati, 2004: 471).

2. Consecuencias en los estimadores en presencia de Autocorrelación.

En presencia de autocorrelación los estimadores de Mínimos Cuadrados Ordinarios siguen distribuidos de forma normal y asintótica, son insesgados, consistentes, pero dejan de ser eficientes (Gujarati, 2004: 471).

A continuación se muestran los resultados de la regresión:

Cuadro 1. Regresión original.

Dependent Variable: CP
Method: Least Squares
Autocorrelación regresión
Sample: 1900Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.60E+08	2.08E+08	3.648706	0.0004
PIB	0.370085	0.048904	7.567576	0.0000
X	11398.62	2709.201	4.207374	0.0001
FBKF	0.584923	0.112519	5.198420	0.0000

R-squared	0.980169	Mean dependent var	4.22E+09
Adjusted R-squared	0.979673	S.D. dependent var	1.09E+09
S.E. of regression	1.55E+08	Akaike info criterion	40.58797
Sum squared resid	2.88E+18	Schwarz criterion	40.67894
Log likelihood	-2512.454	F-statistic	1977.019
Durbin-Watson stat	0.825484	Prob(F-statistic)	0.000000

Recuérdese que para realizar la regresión con EViews, en la barra de comandos escribimos: ls cp c pib x fbkf/ Ok., con lo cual se desplegará una ventana como la contenida en el Cuadro 1.

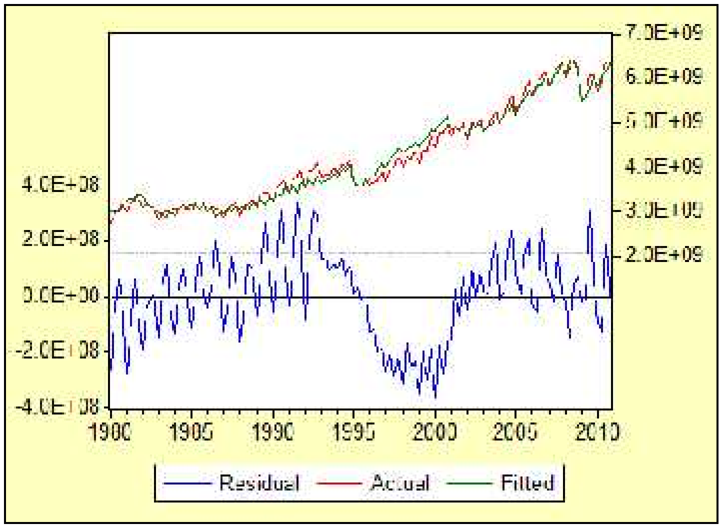
Métodos gráficos para la detección de autocorrelación.

Gráfica de residuales frente al tiempo.

Para determinar si existe o no autocorrelación, se puede ver el comportamiento de los residuales, sí tienen valores del mismo signo seguidos de otro grupo o senda de residuos de signo contrario, es decir, presentan un comportamiento sistemático, hay presencia de autocorrelación. Por otro lado, sí se presentan valores pequeños en los residuos y tienen un comportamiento de forma alternada en cuanto al signo que presentan, es decir, un residuo positivo seguido de un residuo negativo, no existe presencia de autocorrelación.

Con EViews 5 se pueden obtener los residuales de la siguiente manera: en la pantalla de la regresión ir a View/ Actual Fitted, Residual/ Residual Graph, con lo cual se obtendrá una ventana como la mostrada en el Cuadro 2.

Cuadro 2. Valores Reales, Estimados y Residuales.



En el Cuadro 2 se puede observar que los residuales presentan un comportamiento sistemático, es decir tomando valores del mismo signo por un periodo seguido de otros tomando valores del signo contrario, por lo tanto hay presencia de autocorrelación.

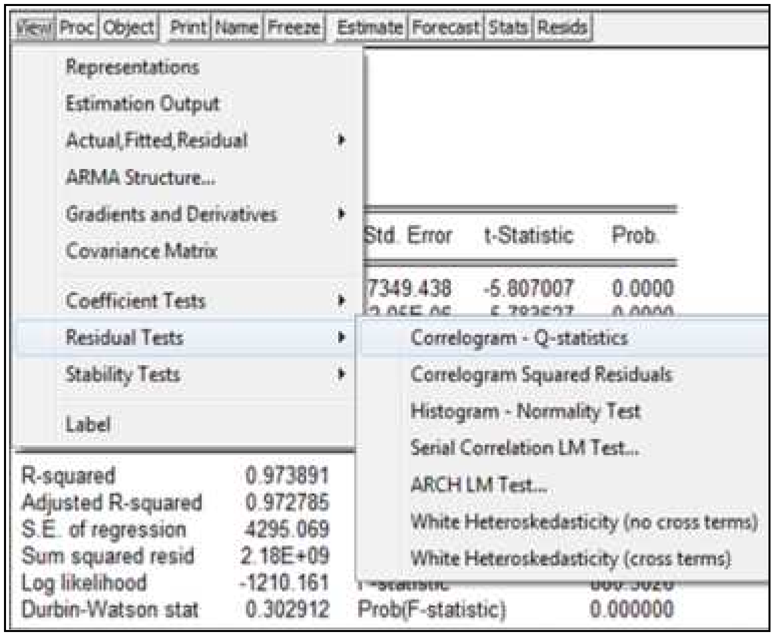
Correlograma.

Es un gráfico de los coeficientes de correlación en función del orden de retardo. Muestra las autocorrelaciones simples o parciales de los residuos en cualquier número especificado para los rezagos correspondientes.

Para ver la prueba en EViews 5 de Correlogram Q Statistics se estima la ecuación de regresión y en la ventana que se obtiene vamos a: View/Residuals Test/ Serial Correlation LM Test/ Lag Specification 30/ Ok.

En el Cuadro siguiente se muestran los pasos para obtener el Correlograma Q con EViews 5.

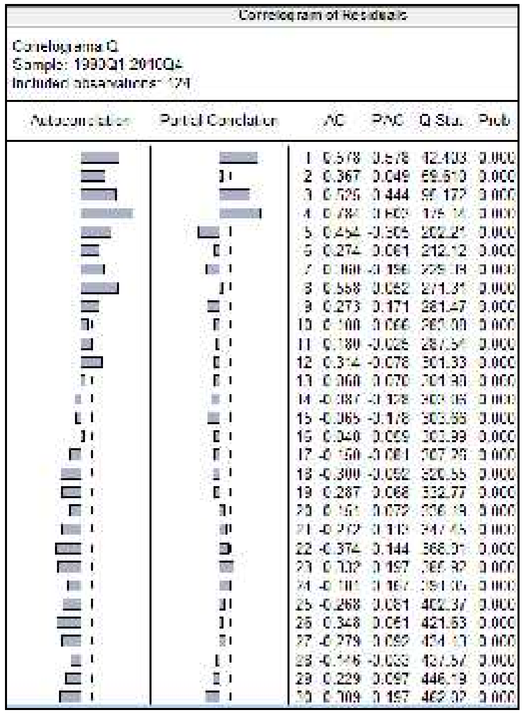
Cuadro 3. Pasos para obtener el Correlograma.



Cuadro 4. Rezagos a incluir.

El Cuadro 4 muestra que para realizar el Correlograma Eviews 5 pide el número de rezagos a incluir (lags to include), en este caso se pondrán 30, este es el número mínimo para poder determinar la presencia o ausencia de autocorrelación con el Correlograma. Los resultados se muestran en el Cuadro siguiente:

Cuadro 5. Correlograma Q.



Correlogram of Residuals

Correlograma Q
Sample: 1980Q1 2010Q4
Included observations: 124

	AC	PAC	Q-Stat	Prob
1	0.578	0.578	42.403	0.000
2	0.367	0.049	59.610	0.000
3	0.525	0.444	[illegible]	0.000
4	[illegible]	[illegible]	[illegible]	0.000
5	0.454	-0.305	202.21	0.000
6	0.274	0.061	212.12	0.000
7	[illegible]	-0.196	[illegible]	0.000
8	0.558	0.052	271.31	0.000
9	0.273	0.171	281.47	0.000
10	0.100	0.066	[illegible]	0.000
11	0.180	-0.025	[illegible]	0.000
12	0.314	-0.078	301.33	0.000
13	0.060	0.070	301.90	0.000
14	-0.087	-0.178	[illegible]	0.000
15	-0.065	-0.178	[illegible]	0.000
16	0.040	0.059	[illegible]	0.000
17	-0.150	-0.081	[illegible]	0.000
18	-0.300	-0.052	320.55	0.000
19	0.287	0.068	332.77	0.000
20	[illegible]	0.072	[illegible]	0.000
21	[illegible]	[illegible]	[illegible]	0.000
22	-0.374	0.144	[illegible]	0.000
23	[illegible]	0.197	[illegible]	0.000
24	[illegible]	[illegible]	[illegible]	0.000
25	-0.268	0.081	402.37	0.000
26	0.348	0.051	421.63	0.000
27	-0.279	[illegible]	[illegible]	0.000
28	[illegible]	[illegible]	[illegible]	0.000
29	0.229	0.097	[illegible]	0.000
30	[illegible]	0.157	[illegible]	0.000

El Cuadro 5 muestra las correlaciones y las correlaciones parciales, las cuales deben de estar dentro de las bandas de confianza para que no exista autocorrelación. Además, otro indicador de la presencia o no de la Autocorrelación es la probabilidad asociada al estadístico Q, la cual debe ser mayor a 0.05 para determinar que no existe autocorrelación. Los valores que asumen los coeficientes de Autocorrelación (AC) y Autocorrelación Parcial (PAC) si tienden a cero significan ausencia de autocorrelación. Asimismo, el valor que asume el estadístico Q debe ser muy pequeño para determinar que no existe Autocorrelación.

En nuestro caso, las correlaciones y correlaciones parciales se encuentran fuera de las bandas de confianza; la probabilidad asociada al estadístico Q es menor a 0.05; en este caso AC y PAC toman valores alejados del cero; y los valores que asume Q son muy grandes. Todo lo anterior indica que el modelo presenta autocorrelación.

Métodos formales para la detección de Autocorrelación.

Anteriormente mencionamos que existen pruebas gráficas y sencillas para tener un primer acercamiento sobre la existencia de la autocorrelación. Pero existen más pruebas para detectar la autocorrelación, estas pruebas son más formales.

1. ***Prueba Durbin –Watson.***

Se utiliza para modelos mayores a 12 datos. Funciona con 2 zonas de indecisión al 95% de confianza, no tiene asociada ninguna distribución de probabilidad y tiene un orden de autocorrelación 1.

El estadístico Durbin-Watson toma valores entre 0 y 4 por lo que sí:

$d \approx 0$ Existe autocorrelación positiva.

$d \approx 2$ No existe autocorrelación.

$d \approx 4$ Existe autocorrelación negativa.

El estadístico *d* de Durbin-Watson depende de: el tamaño de la muestra (n) y el número de regresores (k), por lo que existe una tabla donde pueden encontrarse lo valores que asume (véase anexo).

Se establecieron los límites inferior (d_L) y superior (d_U) para los valores críticos del estadístico *d* en función de las características mencionadas. Si el valor muestra de d es inferior a 2, la verificación o rechazo de la hipótesis es de la siguiente forma:

$d < d_L$ Se rechaza la Ho, existe autocorrelación positiva.

$d_L < d < d_U$. El contraste no es concluyente.

$d_U < d < 2$. No se rechaza la Ho.

Si el valor muestral de *d* es superior a 2, los valores críticos para realizar el contraste son:

$2 < d < 4 - d_u$ No se rechaza la hipótesis nula.

$4 - d_u < d < 4 - d_L$ El contraste no es concluyente.

$4 - d_L < d < 4$ Se rechaza la Ho, existe autocorrelación negativa.

Gráficamente las regiones se verán así:

Cuadro 6. Estadístico Durbin-Watson.

Para realizar la prueba d de Durbin-Watson (DW) primero estimaremos la ecuación como ya se ha realizado anteriormente. En la barra de comandos se escribe la regresión LS CP C PIB FBFK X, con lo cual se obtendrá una ventana como la mostrada en el Cuadro 7. En la parte inferior izquierda se muestra el estadístico DW.

Cuadro 7. Resultados de la estimación.

Dependent Variable: CP
Method: Least Squares
Modelo original: Autocorrelación
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.60E+08	2.08E+08	3.640706	0.0004
PID	0.370085	0.048904	7.567576	0.0000
X	11393.62	2709.201	4.207374	0.0001
FBKF	0.584923	0.112519	5.198420	0.0000

R-squared	0.980169	Mean dependent var	4.22E+09
Adjusted R-squared	0.979673	S.D. dependent var	1.09E+09
S.E. of regression	1.55E+08	Akaike info criterion	40.58797
Sum squared resid	2.88E+18	Schwarz criterion	40.67894
Log likelihood	-2512.454	F-statistic	1977.019
Durbin-Watson stat	0.825484	Prob(F-statistic)	0.000000

La pantalla de regresión anterior muestra el estadístico Durbin-Watson (sombreado en azul) el cual tiene un valor de 0.825484, el cual servirá para hacer la contrastación. Se obtiene los valores del límite inferior (d_L) y superior (d_U), tomando en cuenta que n= 124 y k=4, se busca los valores en la tabla *d* de Durbin-Watson, los cuales son 1.592 y 1.758, respectivamente.

Entonces se tiene que d= 0.825484 < d_L 1.592. Por tanto, se rechaza la hipótesis nula, existe autocorrelación positiva.

2. *Prueba Breusch-Godfrey.*

Esta prueba permite construir una hipótesis alternativa buscando el orden p (coeficiente de autocorrelación) en un proceso autorregresivo (AR) o el orden de un proceso de medias móviles (MA).

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación.

Ha: Autocorrelación.

La regla de decisión es la siguiente:

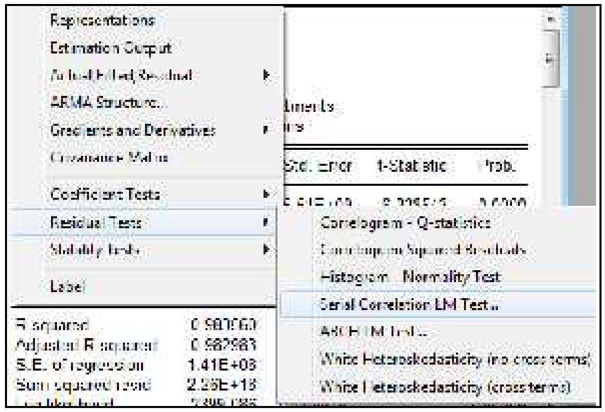
Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

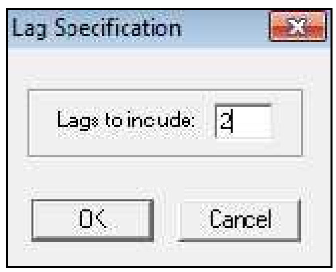
Para realizar la prueba Breusch Godfrey en EViews 5 se estima la ecuación de regresión (ver Cuadro 7); en la ventana de la regresión vamos a: View/Residuals Test/ Serial Correlation LM Test/ Lags to include 2.

En el Cuadro siguiente se muestra cuáles son los pasos para obtener la prueba Breusch-Godfrey con Eviews 5.

Cuadro 8. Pasos para obtener la prueba Breusch-Godfrey.



Cuadro 9. Rezagos a incluir.



Los resultados se muestran a continuación:

Cuadro 10. Prueba Breusch-Godfey.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	59.09053	Probability	0.000000
Obs*R-squared	41.11745	Probability	0.000000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Breusch-Godfrey
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2154430.	1.71E+08	0.012594	0.9900
PIB	-0.004812	0.040147	-0.119864	0.9048
X	244.0403	2224.051	0.109728	0.9128
FBKF	0.033754	0.092465	0.365042	0.7157
RESID(-1)	0.576858	0.075037	7.687622	0.0000

R-squared	0.331834	Mean dependent var	6.62E-07
Adjusted R-squared	0.309375	S.D. dependent var	1.53E+08
S.E. of regression	1.27E+08	Akaike info criterion	40.20088
Sum squared resid	1.93E+18	Schwarz criterion	40.31460
Log likelihood	-2787.454	F-statistic	14.77188
Durbin-Watson stat	2.005426	Prob(F-statistic)	0.000000

Del Cuadro 10 se desprende que la probabilidad asociada al estadístico F es de 0.0000, como es menor a 0.05 se acepta la Ha y concluimos que existe autocorrelación. Asimismo, para determinar que los residuos no están relacionados con ellos mismos, el estimador "RESID(-1)", que son los residuos rezagados un periodo, no deberán ser estadísticamente significativos; mientras que el R^2 deberá ser muy bajo, tendiendo al cero. En este caso se observa que los residuos rezagados si explican a los residuos, por lo que existe autocorrelación en el modelo. Por su parte, el R^2 es de 33.18%, lo cual indica que las variables incluidas en esta regresión auxiliar sí tienen influencia en los residuos, lo cual denota la presencia de autocorrelación en el modelo.

Prueba ARCH LM.

La prueba de Arch LM analiza si la varianza tiene un proceso de comportamiento ARCH, es decir, si depende ó no de ella misma rezagada (n) veces. Además que en un proceso heterocedástico pueda presentarse la varianza del tipo Arch(n).

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación de orden n.

Ha: Autocorrelación de orden n.

La regla de decisión es la siguiente:

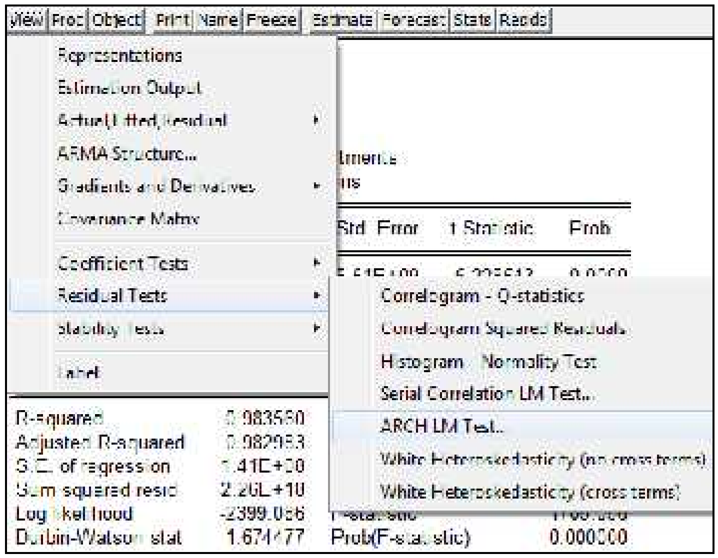
Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

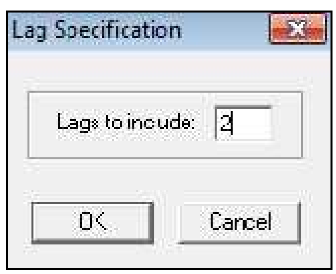
Para generar la prueba Arch LM en EViews 5 se estima la ecuación de regresión (ver Cuadro 7); en la ventana de la regresión vamos a: View/Residuals Test/ ARCH LM Test/ Lags to include 2.

En el Cuadro siguiente se muestra cuales son los pasos para obtener la prueba ARCH LM con Eviews 5:

Cuadro 11. Pasos para obtener la prueba ARCH LM.



Cuadro 12. Rezagos a incluir.



Los resultados son los siguientes:

Cuadro 13. Prueba ARCH LM.

ARCH Test:

F-statistic	7.150879	Probability	0.008528
Obs*R-squared	6.863457	Probability	0.008798

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
ARCH LM Test
Sample (adjusted): 1980Q2 2010Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.74E+16	3.36E+15	5.199842	0.0000
RESID^2(-1)	0.234603	0.087731	2.674111	0.0085

R-squared	0.055800	Mean dependent var	2.29E+16
Adjusted R-squared	0.047997	S.D. dependent var	3.00E+16
S.E. of regression	2.93E+16	Akaike info criterion	78.68545
Sum squared resid	1.04E+35	Schwarz criterion	78.73117
Log likelihood	4837.155	F-statistic	7.150879
Durbin-Watson stat	2.068437	Prob(F-statistic)	0.008528

Del Cuadro 13 se desprende que la probabilidad asociada al estadístico F es de 0.08528, como es menor a 0.05 se acepta la Ha y concluimos que existe autocorrelación. La varianza presenta un proceso de comportamiento ARCH(n), es decir, que depende de ella rezagada (n) veces. Por lo que existe la posibilidad que en el proceso heteroscedástico pueda presentarse una varianza de tipo ARCH (n). Asimismo, obsérvese que la variable "RESID^2(-1)" es estadísticamente significativa, ello indica que ésta explica a la variable dependiente "RESID^2", por lo que, concluimos que existe autocorrelación. Sin embargo, obsérvese que en este caso el R^2 es muy pequeño, a saber 5.58%, lo cual indica que las variables explicativas incluidas en la regresión auxiliar impactan escasamente en el comportamiento de la variable dependiente. Sin embargo, la prueba F de ARCH LM nos indica que existe autocorrelación.

Métodos de Corrección:

1. Cambiar la forma funcional.
2. Método de rezagos distribuidos Koyck. Supone que las β tienen el mismo signo. Postula que cada coeficiente b sucesivo es numéricamente inferior a cada b anterior, lo cual implica que a medida que se retorna a un pasado distante, el efecto de ese rezago sobre Y_t se hace progresivamente menor.
3. El esquema autoregresivo de primer orden de Markov supone que la perturbación en el tiempo actual está linealmente relacionada con el término de perturbación en el tiempo anterior, la medida de esta interdependencia esta dada por el coeficiente de autocorrelación p, que se conoce como el esquema AR (1).
4. Método Iteractivo de Cochrane-Orcutt. Es iteractivo por que hace aproximaciones sucesivas, comenzando con algún valor inicial p. El

objetivo es proporcionar un estimador de (p), que pueda utilizarse para obtener los estimadores de MCG de los parámetros. Puede utilizar no sólo un esquema AR (1) sino esquemas autoregresivos de orden superior.

5. Cuando la autocorrelación es pura se puede utilizar Mínimos Cuadrados Generalizados (MCG).

Propuesta de corrección.

En la primera regresión $CP = f(PIB + FBKF + X$), se encontró que existe autocorrelación entre los datos, es por tal motivo que se reespecifica la regresión para así poder corregir el problema de autocorrelación.

En economía la dependencia de una variable Y con respecto de otra u otras variables X raramente es instantánea, frecuentemente Y responde a X en un lapso de tiempo, el cual se denomina rezago.

Existen razones para utilizar rezagos las cuales son: Razones psicológicas esto se da como resultado de la fuerza de un hábito, la gente no cambia sus hábitos de consumo inmediatamente después de algún cambio. Las razones tecnológicas ocurren cuando se adiciona algún elemento como el capital. Razones institucionales por ejemplo las obligaciones contractuales pueden impedir a las empresas cambiar su fuente de trabajo o materias primas.

El número de rezago determina cual es el dato que se está perdiendo, la correlación de los datos ya no es de uno a uno.

El esquema Autoregresivo AR, es autoregresivo porque es sobre la regresión de residuos, sobre ella misma con un rezago de un periodo. El esquema AR puede combinar con una media móvil (MA) el mecanismo de promedio móvil ARMA.

Entonces de acuerdo a los métodos de corrección antes mencionados se reespecificó la regresión adicionándole logaritmos, rezagos y autoregresivos.

La ecuación de regresión que se propone es:

$CP = f(LogPIB(-2) + LOGFBKF + X)AR(3))$

Dónde:

Log PIB(-2)= es el logaritmo del PIB con el rezago número 2.

Log FBKF= es el logaritmo de la FBKF.

AR (3)= es un autoregresivo de orden 3.

Recordando la prueba de hipótesis para las pruebas formales:

Ho: No Autocorrelación.

Ha: Autocorrelación.

Los criterios de decisión son los siguientes:

Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Estimación de la regresión.

Como se reespecificó la ecuación de regresión se volverá a hacer la estimación en EViews 5. En la barra de comandos escribir la regresión LS CP C LPIB(-2) LFBKF X AR(3), obteniendo la siguiente ventana:

Cuadro 14. Resultados de la estimación.

Dependent Variable: CP
Method: Least Squares
Autocorrelación: corregido.
Sample (adjusted): 1981Q2 2010Q4
Included observations: 119 after adjustments
Convergence achieved after 17 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-3.43E+10	5.51E+09	-6.226540	0.0000
LPIB(-2)	6.73E+08	2.85E+08	2.362730	0.0198
LFBKF	1.09E+09	1.41E+08	7.711856	0.0000
X	21229.98	2139.574	9.922526	0.0000
AR(3)	0.544072	0.078143	6.962077	0.0000

R-squared	0.983560	Mean dependent var	4.27E+09
Adjusted R-squared	0.982983	S.D. dependent var	1.08E+09
S.E. of regression	1.41E+08	Akaike info criterion	40.40480
Sum squared resid	2.26E+18	Schwarz criterion	40.52157
Log likelihood	-2399.086	F-statistic	1705.056
Durbin-Watson stat	1.674477	Prob(F-statistic)	0.000000

Inverted AR Roots	.82	-.41+.71i	-.41-.71i

Del Cuadro 14 se desprende que la ecuación de regresión es:

ĈP = -34335589630.0 + 673359391.7*LPIB(-2) + 1087790759*LFBKF + 21229.98165*X + [AR(3)=0.5440719949]

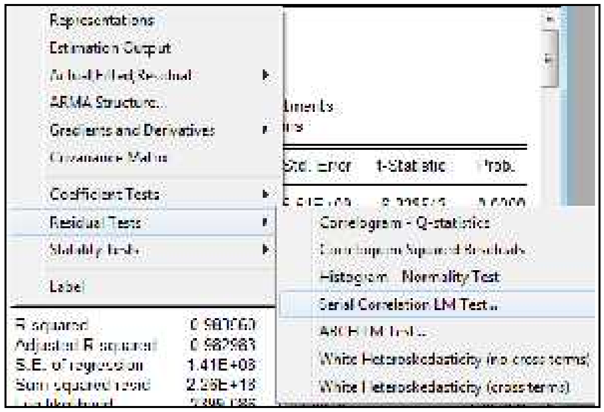
Detección de autocorrelación.

Prueba Breusch-Godfrey.

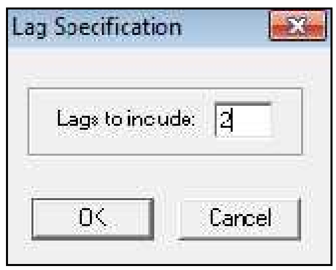
Para ver la prueba en EViews 5 de Breusch-Godfrey, se estima la ecuación de regresión (ver Cuadro 8); en la ventana de la regresión vamos a: View/Residuals Test/ Serial Correlation LM Test/ Lags to include 2.

En el Cuadro siguiente se muestra cuales son los pasos para obtener la prueba Breusch-Godfrey con Eviews 5.

Cuadro 15. Pasos para obtener la prueba Breusch-Godfrey.



Cuadro 16. Rezagos a incluir.



El Cuadro 16 muestra que en Eviews 5 para realizar esta prueba se pide determinar el número de rezagos a incluir (lags to include), en este caso, de acuerdo a como se reespecificó el modelo los rezagos a incluir son 2. Los resultados se muestran a continuación:

Cuadro 17. Prueba Breush- Godrey.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.641215	Probability	0.196969
Obs*R-squared	3.308279	Probability	0.183767

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Breusch-Godfrey
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.59E+09	5.60E+09	0.284037	0.7767
LTIB(-2)	84074322	2.92E+08	0.287833	0.7740
LTPKF	13050105	1.42E+08	0.097509	0.9224
X	-390.7218	2155.066	-0.181304	0.8565
AR(3)	0.019904	0.087073	0.228594	0.8196
RESID(-1)	0.150725	0.089226	1.677324	0.0963
RESID(-2)	0.093284	0.105334	0.881076	0.3910

R-squared	0.023473	Mean dependent var	-0.000110
Adjusted R-squared	-0.020570	S.D. dependent var	1.38E+08
S.E. of regression	1.40E+08	Akaike info criterion	40.40953
Sum squared resid	2.20E+18	Schwarz criterion	40.57301
Log likelihood	-2397.367	F-statistic	0.547072
Durbin-Watson stat	1.990010	Prob(F-statistic)	0.771334

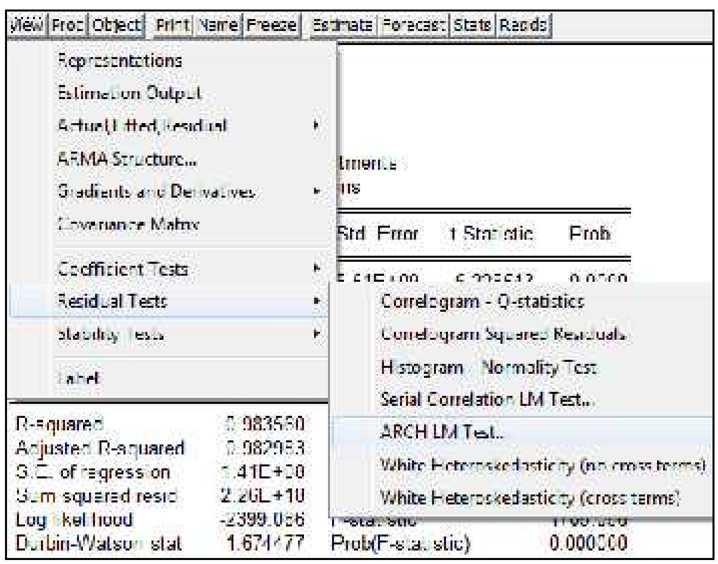
En el Cuadro 17 se muestra la prueba Breusch-Godfrey, en la cual la probabilidad asociada del estadístico F es de 0.198369, lo cual es mayor a 0.05 entonces se acepta la hipótesis de no autocorrelación. Al incluir dos rezagos en el PIB (-2), quiere decir que los primeros dos datos se están perdiendo, por tanto, ya no será del primer dato con el primer dato. Los estimadores siguen teniendo las propiedades de eficiencia e insesgadez. Nótese también que todas las variables independientes incluidas en esta regresión auxiliar son no estadísticamente significativas, por lo que no tienen ninguna relación con la variable dependiente, en este caso los residuales. Asimismo, el estadístico F del modelo indica que en su conjunto las variables independientes no explican a la dependiente. En cuanto al R^2 obtenido, éste asciende a 2.84%, lo que indica que las variables independientes incluidas en la regresión auxiliar no explican los cambios en los residuales. Todo ello es indicativo de que no existe autocorrelación en el modelo.

Prueba Arch LM.

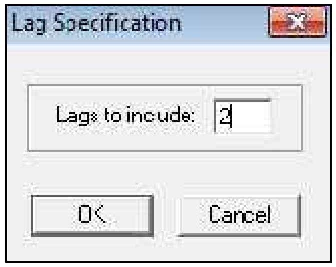
Para ver la prueba en EViews 5 de Arch LM Test, se estima la ecuación de regresión; en la ventana de la regresión vamos a: View/Residuals Test/ ARCH LM Test/ Lags to include 2.

En el Cuadro siguiente se muestra cuáles son los pasos para obtener la prueba ARCH LM con Eviews 5.

Cuadro 18. Pasos para obtener la prueba ARCH-LM.



Cuadro 19. Rezagos a incluir.



El Cuadro 19 muestra que en Eviews 5 para realizar esta prueba se pide determinar el número de rezagos a incluir (lags to include), en este caso, de acuerdo a como se reespecificó el modelo los rezagos a incluir son 2. Los resultados se muestran a continuación:

Cuadro 20. Prueba ARCH LM.

ARCH Test:

F-statistic	1.435969	Probability	0.242154
Obs*R-squared	2.875086	Probability	0.237511

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
Arch LM test
Sample (adjusted): 1981Q4 2010Q4
Included observations: 117 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.00E+16	3.55E+15	5.640463	0.0000
RESID^2(-1)	-0.133721	0.093465	-1.430702	0.1553
RESID^2(-2)	0.063725	0.092433	0.689417	0.4920

R-squared	0.024573	Mean dependent var	1.07E+16
Adjusted R-squared	0.007461	S.D. dependent var	2.56E+16
S.E. of regression	2.55E+16	Akaike info criterion	76.41596
Sum squared resid	7.40E+34	Schwarz criterion	76.48579
Log likelihood	-4584.334	F-statistic	1.435969
Durbin-Watson stat	1.995160	Prob(F-statistic)	0.242154

Del Cuadro 20 se observa que se aplicó la prueba de Arch LM la cual tiene una probabilidad asociada al estadístico F de 0.242154 que es mayor a 0.05, por lo tanto se acepta la hipótesis nula Ho: no existe correlación. Además que la varianza no presenta un proceso de comportamiento ARCH(n), es decir no depende de ella misma rezagada en el tiempo, se descarta la posibilidad de que exista un proceso heterocedástico que pueda presentar la varianza del tipo ARCH(n).

Correlograma.

Para ver la prueba en EViews 5 de Correlogram Q Statistics se estima la ecuación de regresión; en la ventana de la regresión vamos a: View/Residuals Test/

Correlogram Q Statistics/ Lags to include 30, con lo cual se mostrará una ventana como la contenida en el Cuadro 21.

Cuadro 21. Correlograma.

Correlogram of Residuals

Correlograma Q
Sample: 1981Q2 2010Q4
Included observations: 119
Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.133	0.138	2.3323	
		2	0.027	0.047	2.4218	0.120
		3	0.016	0.027	2.4561	0.293
		4	0.590	0.595	45.977	0.000
		5	0.011	0.236	45.992	0.000
		6	0.025	0.072	46.070	0.000
		7	0.059	0.073	46.522	0.000
		8	0.424	0.150	69.869	0.000
		9	0.012	0.044	69.888	0.000
		10	0.013	0.056	69.911	0.000
		11	0.050	0.023	70.245	0.000
		12	0.357	0.070	86.895	0.000
		13	0.020	0.023	86.950	0.000
		14	0.010	0.054	86.964	0.000
		15	0.154	0.130	90.260	0.000
		16	0.184	0.057	95.012	0.000
		17	0.129	0.123	97.345	0.000
		18	0.134	0.160	99.892	0.000
		19	0.272	0.120	110.56	0.000
		20	0.069	0.070	111.26	0.000
		21	0.147	0.036	114.41	0.000
		22	0.107	0.011	116.11	0.000
		23	0.220	0.026	123.30	0.000
		24	0.074	0.009	124.21	0.000
		25	0.120	0.003	126.47	0.000
		26	0.067	0.001	127.63	0.000
		27	0.166	0.007	131.92	0.000
		28	0.050	0.002	132.32	0.000
		29	0.105	0.003	134.14	0.000

El correlograma muestra que tanto en la columna de Autocorrelación (AC) como en correlación parcial (PAC) algunas barras se salen de las bandas de confianza. Asimismo, los primeros valores de AC y PAC son relativamente altos. Por lo que respecta a los valores del estadístico Q, como puede apreciarse éstos son muy grandes y sus probabilidades asociadas son menores a 0.05. Todo lo anterior es indicativo de que aún persiste la autocorrelación en el modelo.

Sin embargo, si bien el correlograma nos indica que aún persiste la autocorrelación en el modelo, las pruebas Breusch-Godfrey (BG) y ARCH LM nos indican lo contrario. Para fines prácticos nos quedaremos con los resultados de las pruebas BG y ARCH LM ya que, por construcción, son pruebas más robustas y por ende, más confiables. En este sentido, concluimos que las operaciones realizadas fueron idóneas para resolver el problema de la autocorrelación.

Ejemplo 2: Modelo de exportaciones.

En este ejemplo se buscará hacer una regresión donde las Exportaciones (X) estén explicadas por el Producto Interno Bruto (PIB), la Formación Bruta de Capital Fijo (FBKF), el Tipo de Cambio (TC) el cual suponemos que afecta de manera negativa a las exportaciones y finalmente el Consumo de Gobierno (CG).

$$X = F(PIB + FBKF + INPC - TC + CG)$$

La prueba de hipótesis será la siguiente.

Ho: No autocorrelación.

Ha: Autocorrelación.

Primero se estima la ecuación de regresión. En la barra de comandos escribir la regresión LS X C PIB FBKF INPC TC CG, obteniendo la siguiente ventana:

Cuadro 22. Regresión original.

Dependent Variable: X
Method: Least Squares
Autocorrelación.
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-42678.24	7319.438	-5.807007	0.0000
PIB	1.19E-05	2.05E-06	5.783627	0.0000
FBKF	-1.91E-06	3.80E-06	-0.501351	0.6171
INPC	653.3958	95.65518	6.830741	0.0000
TC	-2900.617	462.6962	-6.268944	0.0000
CG	-6.03E-06	4.11E-06	-1.466345	0.1452

R-squared	0.973891	Mean dependent var	33392.56
Adjusted R squared	0.972785	S.D. dependent var	26035.35
S.E. of regression	4295.069	Akaike info criterion	19.61550
Sum squared resid	2.10E+09	Schwarz criterion	19.75197
Log likelihood	-1210.161	F-statistic	880.3026
Durbin-Watson stat	0.302912	Prob(F-statistic)	0.000000

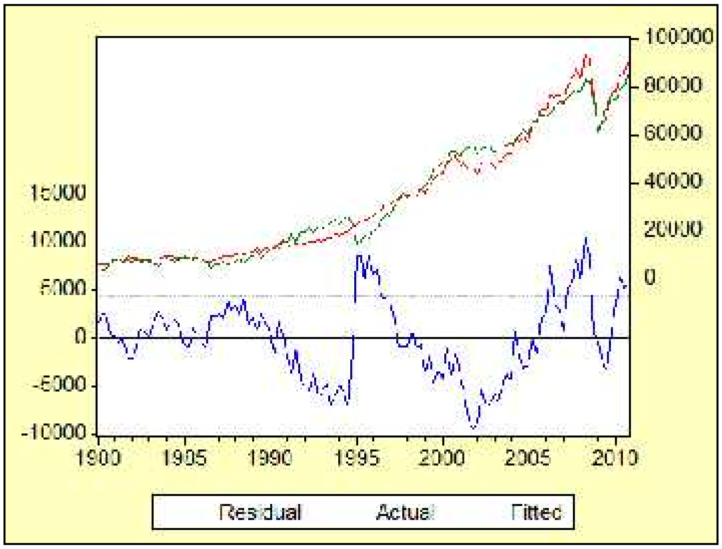
A continuación procedemos a aplicar tanto los métodos no formales como formales para la detección de la autocorrelación en el modelo.

Métodos gráficos para la detección de la autocorrelación.

Gráfica de residuales frente al tiempo.

Con EViews 5 se pueden obtener los residuales de la siguiente manera: en la pantalla de la regresión ir a View/ Actual Fitted, Residual/ Residual Graph, con lo cual se obtendrá una ventana como la que se muestra en el Cuadro siguiente:

Cuadro 23. Valores Reales, Estimados y Residuales.



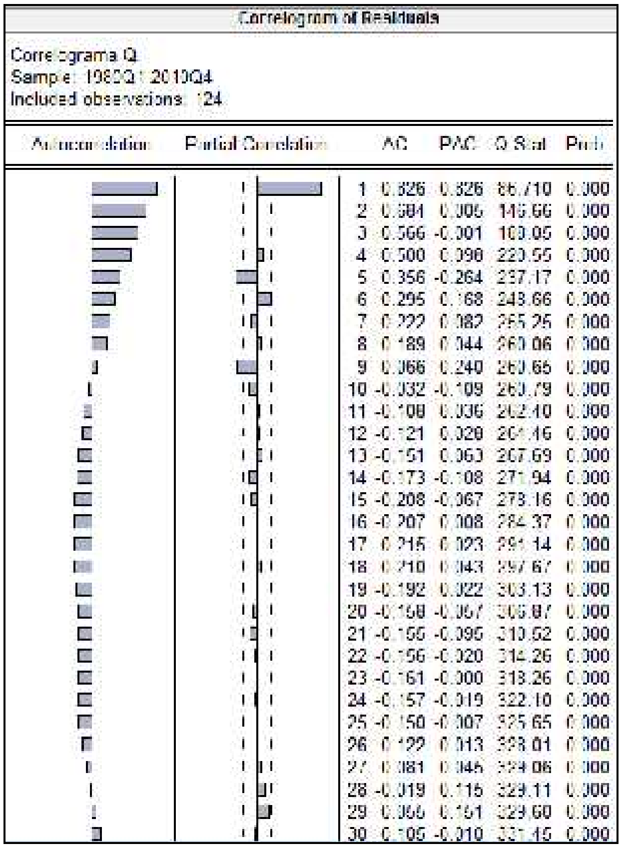
En el Cuadro 23 se puede observar que los residuales presentan un comportamiento sistemático, es decir, tomando valores del mismo signo por un periodo seguido de otro periodo tomando valores del signo contrario, por lo tanto, a partir de la gráfica de residuales se puede concluir que el modelo presenta autocorrelación.

Correlograma.

Es un gráfico de los coeficientes de autocorrelación en función del orden de retardo. Muestra las autocorrelaciones simples o parciales de los residuos en cualquier número especificados para los rezagos correspondientes.

Para ver la prueba en EViews 5 de Correlogram Q Statistics se estima la ecuación de regresión; en la ventana de la regresión vamos a: View/Residuals Test/Correlograma Q Statistics/ Lag Specification 30 (ver ejercicio 1 de autocorrelación), con lo cual se obtiene el siguiente Cuadro:

Cuadro 24. Correlograma.



Correlogram of Residuals

Correlograma Q
Sample: 1980Q1 2010Q4
Included observations: 124

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.826	0.826	86.710	0.000
		2	0.684	0.005	146.66	0.000
		3	0.566	-0.001	188.05	0.000
		4	0.500	0.098	220.55	0.000
		5	0.356	-0.264	237.17	0.000
		6	0.295	0.168	248.66	0.000
		7	0.222	0.082	255.25	0.000
		8	0.189	0.044	260.06	0.000
		9	0.066	0.240	260.65	0.000
		10	-0.032	-0.109	260.79	0.000
		11	-0.108	0.036	262.40	0.000
		12	-0.121	0.028	264.16	0.000
		13	-0.151	0.063	267.69	0.000
		14	-0.173	-0.108	271.94	0.000
		15	-0.208	-0.067	278.16	0.000
		16	-0.207	0.008	284.37	0.000
		17	-0.215	0.023	291.14	0.000
		18	-0.210	0.043	297.67	0.000
		19	-0.192	0.022	303.13	0.000
		20	-0.158	-0.057	306.87	0.000
		21	-0.155	-0.095	310.52	0.000
		22	-0.156	-0.020	314.26	0.000
		23	-0.161	-0.000	318.26	0.000
		24	-0.157	-0.019	322.10	0.000
		25	-0.150	-0.007	325.65	0.000
		26	-0.122	0.013	328.01	0.000
		27	-0.081	0.046	329.06	0.000
		28	-0.019	0.116	329.11	0.000
		29	0.066	0.161	329.60	0.000
		30	0.105	-0.010	331.45	0.000

En el Cuadro 24 se puede observar que las autocorrelaciones (AC) y las correlaciones parciales (PAC) están fuera de las bandas de confianza. Asimismo, los valores que éstas asumen son altos. En cuanto a los valores del estadístico Q, éstos son muy grandes y sus probabilidades asociadas son menores a 0.05. Todo lo anterior indica que el modelo presenta autocorrelación.

Métodos formales para la detección de la Autocorrelación.

Prueba Durbin–Watson.

Primero estimaremos la ecuación como ya se ha realizado anteriormente. En la barra de comandos se escribe la regresión LS X C PIB FBKF INPC TC CG, con lo cual se obtiene el Cuadro siguiente:

Cuadro 25. Resultados de la Regresión Original.

Dependent Variable: X
Method: Least Squares
Autocorrelación.
Sample: 1980Q1 2010Q4
Included observations: 124

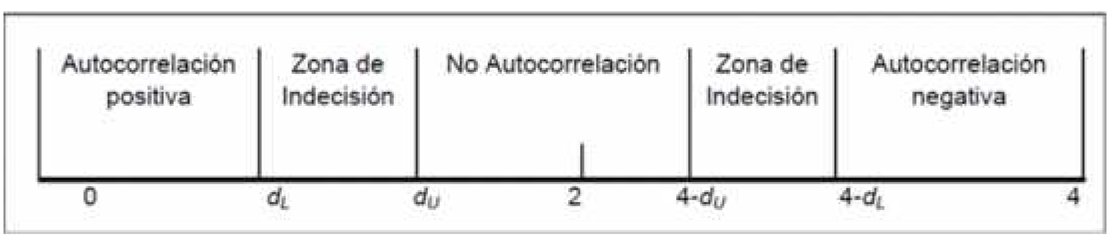
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-42678.24	7319.438	-5.807007	0.0000
PIB	1.19E-05	2.05E-06	5.783627	0.0000
FBKF	-1.91E-06	3.80E-06	-0.501351	0.6171
INPC	653.3958	95.65518	6.830741	0.0000
TC	-2900.617	462.6962	-6.268944	0.0000
CC	-6.03E-06	4.11E-06	-1.466345	0.1452

R-squared	0.973891	Mean dependent var	33392.56
Adjusted R-squared	0.972785	S.D. dependent var	26035.35
S.E. of regression	4295.069	Akaike info criterion	19.61550
Sum squared resid	2.18E+09	Schwarz criterion	19.75197
Log likelihood	-1210.161	F-statistic	880.3026
Durbin-Watson stat	0.302912	Prob(F-statistic)	0.000000

La pantalla de regresión anterior muestra el estadístico Durbin-Watson, mismo que asume un valor de 0.302912. Para poder determinar si existe autocorrelación o no se deben obtener los valores de los límites inferior (d_L) y superior (d_U), tomando en cuenta que n= 124 y k=6, se buscan los valores en la tabla *d* de Durbin-Watson los cuales son 1.651 y 1.817, respectivamente.

Entonces se tiene que $d = 0.302912 < d_L = 1.651$. Por lo que se rechaza la hipótesis nula y concluimos, mediante esta prueba, que existe autocorrelación positiva en el modelo.

Cuadro 26. Durbin-Watson zonas de decisión.



Prueba Breusch-Godfrey.

Esta prueba permite construir una hipótesis alternativa buscando el orden p (coeficiente de autocorrelación) en un proceso autoregresivo o el orden de un proceso de medias móviles (MA).

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación.

Ha: Autocorrelación.

La regla de decisión es la siguiente:

Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Para realizar la prueba en EViews 5 de Breusch-Godfrey, se estima la ecuación de regresión (ver Cuadro 1.1); en la ventana de la regresión vamos a: View/Residuals Test/ Serial Correlation LM Test/ Lags to include 0 (incluir 3 ó 4 rezagos, esto debido a que la prueba Durbin Watson prueba autocorrelación de orden 1, el correlograma de orden 2, además que los datos son trimestrales. Por tanto, será preciso probar Autocorrelación de orden 3 y 4).

Cuadro 27. Prueba Breusch-Godfrey con 3 rezagos.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	100.6473	Probability	0.000000
Obs*R-squared	83.79859	Probability	0.000000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Breusch-Godfrey
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1302.554	3979.247	0.327337	0.7440
PIB	8.21E-07	1.13E-06	0.725798	0.4694
FBKF	7.24E-07	2.12E-06	0.342080	0.7329
INPC	16.81390	50.93425	0.330110	0.7419
TC	20.37289	250.6542	0.081279	0.9354
CG	2.89E-06	2.27E-06	1.274929	0.2049
RESID(-1)	0.869929	0.096478	9.016853	0.0000
RESID(-2)	0.066969	0.124099	0.539640	0.5906
RESID(-3)	-0.095738	0.096188	-0.995841	0.3214

R-squared	0.724152	Mean dependent var	2.49E-11
Adjusted R-squared	0.704995	S.D. dependent var	4206.865
S.E. of regression	2284.933	Akaike info criterion	18.37587
Sum squared resid	6.00E+08	Schwarz criterion	18.58057
Log likelihood	-1130.304	F-statistic	37.74273
Durbin-Watson stat	1.951400	Prob(F-statistic)	0.000000

Del Cuadro 27 se desprende que la probabilidad asociada al estadístico F es de 0.0000, como es menor a 0.05 se acepta la Ha y concluimos que existe autocorrelación entre los términos de error. Asimismo, obsérvese que la variable “Resid(-1)” en la regresión auxiliar (Prueba BG) es estadísticamente significativa, lo que indica que ésta explica a la variable dependiente “Resid”; también, el estadístico F es significativo, indicando que las variables independientes explican, en su conjunto, a la variable dependiente; por su parte, el valor del R^2 es de 72.41%, lo cual indica que en esta proporción las variables independientes explican los cambios en la dependiente. Todo lo anterior es indicativo de que el modelo presenta autocorrelación.

Prueba ARCH LM.

La prueba de ARCH LM analiza si la varianza tiene un proceso de comportamiento ARCH, es decir, si depende o no de ella misma rezagada (n) veces. Además que en un proceso heterocedástico pueda presentarse la varianza del tipo ARCH(n).

Para realizar la prueba en EViews 5 de ARCH LM Test, se estima la ecuación de regresión, en la ventana de la regresión vamos a: View/Residuals Test/ ARCH LM Test/ Lags to include 0 (incluir 3 ó 4 rezagos esto debido a que la prueba Durbin Watson prueba autocorrelación de orden 1, el correlogramoa de orden 2, además de que los datos son trimestrales, por lo que será preciso probar Autocorrelación de orden 3 y 4 mediante esta prueba).

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación de orden n.

Ha: Autocorrelación de orden n.

La regla de decisión es la siguiente:

Si la probabilidad asociada al estadístico F $\geq$ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Cuadro 28. Prueba ARCH LM Test con 3 rezagos.

ARCH Test:

F-statistic	28.90906	Probability	0.000000
Obs*R-squared	51.51001	Probability	0.000000

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
ARCH LM Test
Sample (adjusted): 1980Q4 2010Q4
Included observations: 121 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5180122.	2118714.	2.110941	0.0175
RESID^2(-1)	0.505966	0.092305	5.481459	0.0000
RESID^2(-2)	0.153233	0.102587	1.493682	0.1380
RESID^2(-3)	0.062880	0.092553	0.679491	0.4982

R-squared	0.425703	Mean dependent var	17363231
Adjusted R-squared	0.410977	S.D. dependent var	22922794
S.E. of regression	17592740	Akaike info criterion	36.23637
Sum squared resid	3.62E+16	Schwarz criterion	36.32879
Log likelihood	2188.300	F-statistic	28.90906
Durbin-Watson stat	1.986116	Prob(F-statistic)	0.000000

El Cuadro 28 muestra que la probabilidad asociada al estadístico F es de 0.000, como es menor que 0.05, entonces se acepta Ha, concluyendo que existe autocorrelación en el modelo. En este sentido, la varianza presenta un proceso de comportamiento ARCH(n), depende de ella rezagada (n) veces. Por lo que existe la posibilidad de que en el proceso heterocedástico pueda presentarse una varianza de tipo ARCH (n). La existencia de autocorrelación en el modelo se corrobora con el hecho de que existen variables independientes en la regresión auxiliar (ARCH LM) que son estadísticamente significativas, indicando que, en este caso, “RESID^2(-1)” explica a la variable dependiente “RESID^2”; el estadístico F del modelo indica que en su conjunto las variables independientes explican a la dependiente; y el R^2 de la regresión auxiliar es de 42.57%.

Propuesta de corrección.

De acuerdo con los métodos de corrección anteriormente mencionados se propone una nueva regresión, pues la regresión X= F (PIB+FBKF+INPC-TC+CG) presentó problemas de autocorrelación. Esta nueva regresión incluye a la variable dependiente explicada por ella misma rezagada en el tiempo un periodo, lo cual puede suceder debido a la naturaleza de la autocorrelación pura de los datos.

La regresión reespecificada es la siguiente:

$$LX = F(LPIB(-2) + INPC(-3) - TC(-3) + LX(-1)$$

Dónde:

LPIB(-2)= es el logaritmo del PIB rezagado 2 periodos.

LINPC(-3)= es el logaritmo del INPC rezagado 3 periodos.

TC(-3)= es el TC rezagado 3 periodos.

LX= es el logaritmo de las X rezagadas un periodo.

Estimación de la regresión.

Como se reespecificó el modelo se volverá a hacer la estimación en EViews 5. Así, en la barra de comandos escribiremos la regresión de la forma siguiente: LS LX C LPIB(-2) INPC(-3) TC(-3) LX(-1), con lo cual se obtendrá una ventana como la que se muestra en el siguiente Cuadro:

Cuadro 29. Resultados de la regresión reespecificada.

Dependent Variable: LX
Method: Least Squares
Date: 10/10/11 Time: 22:01
Autocorrelación corrección
Included observations: 121 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.634343	3.453056	-2.790092	0.0062
LPIB(-2)	0.535423	0.168714	3.173553	0.0019
INPC(-3)	-0.002250	0.001079	-2.086179	0.0392
TC(-3)	0.037094	0.007757	4.781765	0.0000
LX(-1)	0.747021	0.053655	13.92278	0.0000

R-squared	0.994315	Mean dependent var	10.10721
Adjusted R-squared	0.994119	S.D. dependent var	0.346670
S.E. of regression	0.064931	Akaike info criterion	-2.590544
Sum squared resid	0.489057	Schwarz criterion	-2.475016
Log likelihood	161.7279	F-statistic	5071.913
Durbin-Watson stat	1.900413	Prob(F-statistic)	0.000000

Del Cuadro 29 se desprende que la ecuación de regresión es:

*LX = -9.634343253 + 0.5354227092*LPIB(-2) - 0.002250051676*INPC(-3) + 0.03709384495*TC(-3) + 0.7470208011*LX(-1)*

A continuación se procederá a determinar si con la reespecificación del modelo se corrigió el problema de la Autocorrelación.

Detección de autocorrelación.

Prueba Breusch-Godfrey.

Para obtener la prueba en EViews 5 de Breusch-Godfrey, se estima la ecuación de regresión; en la ventana de la regresión vamos a: View/Residuals Test/ Serial

Correlation LM Test/ Lags to include 3.

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación.

Ha: Autocorrelación.

La regla de decisión es la siguiente:

Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Cuadro 30. Corrección prueba Breusch-Godfrey.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	0.208296	Probability	0.890476
Obs*R-squared	0.665448	Probability	0.881299

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Breusch-Godfrey
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.569419	4.564624	-0.343822	0.7316
LPIB(-2)	0.084148	0.233657	0.360135	0.7194
INPC(-3)	-0.000118	0.001125	-0.104499	0.9170
TC(-3)	0.003012	0.010450	0.288210	0.7737
LX(-1)	-0.033844	0.089882	-0.376538	0.7072
RESID(-1)	0.087603	0.143699	0.609624	0.5433
RESID(-2)	0.026449	0.101806	0.259800	0.7955
RESID(-3)	-0.024943	0.108147	-0.230640	0.8180

R-squared	0.005500	Mean dependent var	3.50E-16
Adjusted R-squared	-0.056107	S.D. dependent var	0.063839
S.E. of regression	0.065606	Akaike info criterion	-2.546472
Sum squared resid	0.486368	Schwarz criterion	-2.361627
Log likelihood	162.0616	F-statistic	0.089270
Durbin-Watson stat	1.977523	Prob(F-statistic)	0.998775

En el Cuadro 30 se observa que la probabilidad asociada al estadístico F es de 0.890476, como es mayor a 0.05, se acepta la Ho y concluimos que el modelo no presenta autocorrelación. Asimismo, nótese que ninguna variable independiente de la regresión auxiliar (prueba BG) es estadísticamente significativa; la estadística F de la regresión auxiliar posee una probabilidad mayor a 0.05, indicando que las variables independientes en su conjunto no explican a la dependiente, en este caso, a los residuales; y, el valor del R^2 es ínfimo, a saber igual a 0.5%. Los resultados anteriores indican que el modelo ha quedado libre de autocorrelación.

Prueba ARCH-LM.

Para realizar la prueba en EViews 5 de Arch LM Test, se estima la ecuación de regresión (ver Cuadro 20); en la ventana de la regresión vamos a: View/Residuals Test/ ARCH LM Test/ Lags to include 3.

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación de orden n.

Ha: Autocorrelación de orden n.

La regla de decisión es la siguiente:

Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Cuadro 31. Corrección prueba ARCH LM.

ARCH Test:

F-statistic	2.646785	Probability	0.052416
Obs*R-squared	7.683773	Probability	0.053020

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
ARCH-LM Test
Sample (adjusted): 1981Q3 2010Q4
Included observations: 118 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.002803	0.000800	3.504702	0.0007
RESID^2(-1)	0.096490	0.093292	1.034279	0.3032
RESID^2(-2)	0.215823	0.084848	2.543651	0.0123
RESID^2(-3)	-0.046230	0.087086	-0.530855	0.5966

R-squared	0.065117	Mean dependent var	0.003872
Adjusted R-squared	0.040515	S.D. dependent var	0.006534
S.E. of regression	0.005400	Akaike info criterion	-7.231690
Sum squared resid	0.004670	Schwarz criterion	-7.137769
Log likelihood	430.6697	F-statistic	2.646785
Durbin-Watson stat	1.872463	Prob(F-statistic)	0.052416

En el Cuadro 31 observamos que la probabilidad asociada al estadístico F es de 0.052416, como es mayor a 0.05, se acepta la Ho y concluimos que el modelo no presenta autocorrelación. Asimismo, el estadístico F de la regresión auxiliar (prueba ARCH LM) indica que las variables independientes en su conjunto no explica a la variable “RESID^2”; y el valor del R^2 asciende a 6.51%, lo cual indica que el 6.51% de los cambios en los residuales al cuadrado se explican por ellos mismos rezagadaos 1, 2 y 3 periodos. Como el valor que asume el R^2 es muy bajo, se concluye que el modelo no presenta autocorrelación.

Correlograma.

Para realizar la prueba en EViews 5 de Correlogram Q Statistics se estima la ecuación de regresión; en la ventana de la regresión vamos a: View/Residuals Test/ Correlograma Q Statistics/ Lags to include 30, con lo que se obtendrá un Cuadro como el siguiente:

Cuadro 32. Correlograma.

Correlogram of Residuals

Correlograma Q
Sample: 1980Q4 2010Q4
Included observations: 121

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.046	0.045	0.2666	0.613
		2	0.015	0.013	0.2820	0.868
		3	-0.036	-0.037	0.4440	0.931
		4	0.076	0.080	1.1870	0.880
		5	-0.177	-0.185	5.1855	0.394
		6	0.060	0.080	5.6465	0.464
		7	-0.078	-0.082	6.4358	0.490
		8	-0.015	-0.021	6.4649	0.596
		9	0.102	0.068	7.8346	0.551
		10	0.019	-0.019	7.8825	0.640
		11	0.113	0.082	9.5981	0.567
		12	0.110	0.095	11.256	0.507
		13	0.018	0.017	11.300	0.586
		14	0.065	0.028	11.889	0.615
		15	-0.034	-0.012	12.050	0.675
		16	0.033	0.025	12.203	0.730
		17	-0.141	-0.107	15.066	0.591
		18	0.086	0.079	16.138	0.583
		19	-0.111	-0.114	17.941	0.526
		20	0.072	0.076	18.709	0.541

En el Cuadro 32 se puede ver que las correlaciones (AC) y las correlaciones parciales (PAC) están dentro de las bandas de confianza, por lo que se determina que no existe Autocorrelación. Asimismo, se observa que los valores que asumen los coeficientes de Autocorrelación (AC) y Autocorrelación Parcial (PAC) son tendientes a cero, lo cual indica la ausencia de autocorrelación. Por lo que respecta a los valores que asume el estadístico “Q”, observamos que estos son pequeños y tienden a cero, lo cual se corrobora con su probabilidad asociada, la cual debe ser mayor o igual a 0.05 para indicar la ausencia de Autocorrelación, en este caso, mediante el uso de dicho estadístico se concluye que no existe Autocorrelación.

En general, se puede concluir que se resolvió el problema de la autocorrelación entre los términos de error. Dicha solución se obtuvo mediante la reespecificación del modelo, para ello se realizaron transformaciones en las variables, a través de su conversión a logaritmos e incluyendo a la variable dependiente como independiente pero rezagada en el tiempo.

La regresión finalmente no presenta autocorrelación, las varianzas y los errores estándar son mínimos y los estimadores recuperaron la propiedad de la eficiencia. Como la regresión cuenta con 3 rezagos, esto quiere decir, que la correlación de los datos no es de uno a uno, entonces se pierden los 3 primeros datos.

Ejemplo 3: Modelo de Importaciones.

El siguiente ejercicio pretende explicar las Importaciones (M) en función del Consumo Privado (CP), la Formación Bruta de Capital Fijo (FBKF) y el Consumo de Gobierno (CG).

$$M - f\ (CP + FBKF + CG)$$

La prueba de hipótesis será la siguiente:

Ho: No autocorrelación.

Ha: Autocorrelación.

Primero se estima la ecuación de regresión. Así, en la barra de comandos de EViews escribimos la regresión: LS M CP FBKF CG. Los resultados se muestran a continuación:

Cuadro 33. Regresión Original.

Dependent Variable: M
Method: Least Squares
Autocorrelación
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-61173.23	2205.992	-27.73048	0.0000
CP	1.98E-05	1.35E-06	14.69184	0.0000
FBKF	1.14E-05	3.30E-06	3.454748	0.0008
CG	-8.37E-07	3.47E-06	-0.255585	0.7987

R-squared	0.976706	Mean dependent var	35639.22
Adjusted R-squared	0.976123	S.D. dependent var	26380.26
S.E. of regression	4076.775	Akaike info criterion	19.49573
Sum squared resid	1.99E+09	Schwarz criterion	19.58670
Log likelihood	-1204.735	F-statistic	1677.146
Durbin-Watson stat	0.738135	Prob(F-statistic)	0.000000

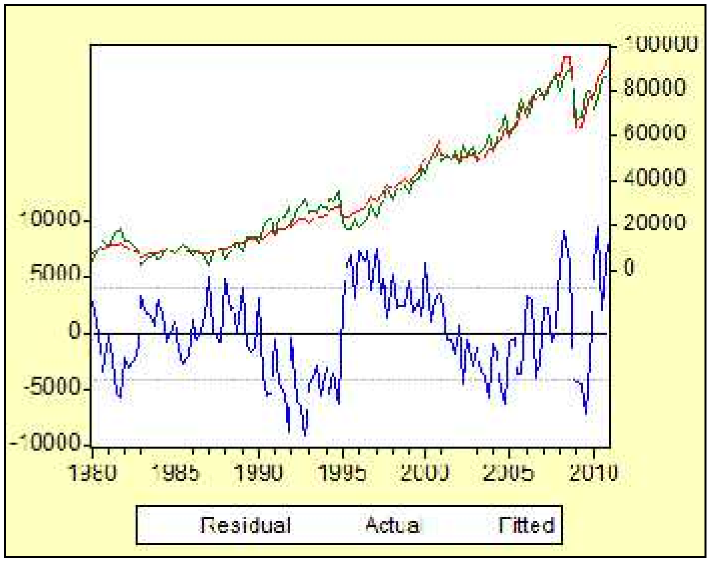
A continuación procedemos a aplicar tanto los métodos no formales como formales para la detección de la autocorrelación en el modelo.

Métodos gráficos para la detección de la autocorrelación.

Gráfica de residuales frente al tiempo.

Con EViews 5 se pueden obtener los residuales de la siguiente manera: en la pantalla de la regresión ir a View/ Actual Fitted, Residual/ Residual Graph, con lo cual se obtiene un Cuadro como el siguiente:

Cuadro 34. Valores Reales, Estimados y Residuales.



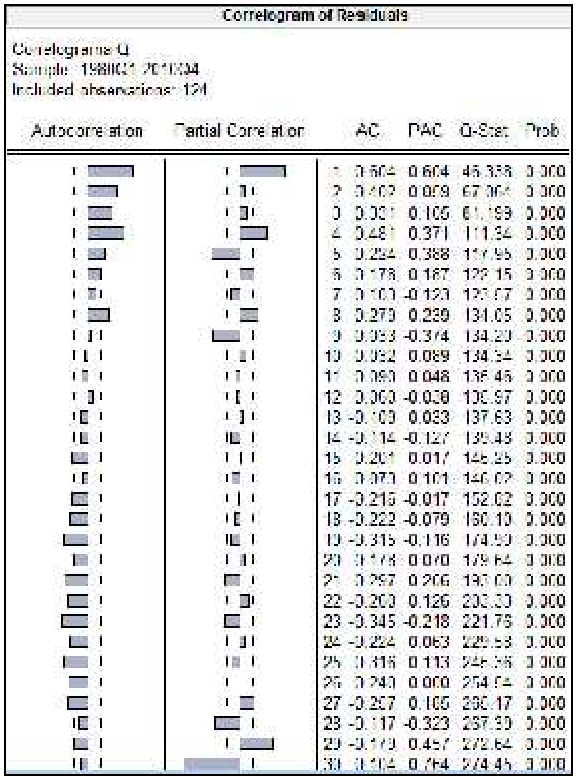
A partir del Cuadro anterior se observa que los residuales presentan un comportamiento sistemático, es decir tomando valores del mismo signo por un periodo seguido de otro periodo tomando valores del signo contrario, por lo tanto el gráfico indica presencia de autocorrelación en el modelo.

Correlograma.

Es un gráfico de los coeficientes de autocorrelación en función del orden de retardo. Muestra las autocorrelaciones simples o parciales de los residuos en cualquier número especificado para los rezagos correspondientes.

Para obtener la prueba en EViews 5 de Correlogram Q Statistics se estima la ecuación de regresión; en la ventana de la regresión vamos a: View/Residuals Test/ Serial Correlation LM Test/ Lag Specification 30 (ver ejercicio 1 de autocorrelación), con lo cual se obtienen un Cuadro como el siguiente:

Cuadro 35. Correlograma.



Correlogram of Residuals

Correlograma Q
Sample: 1980Q1 2010Q4
Included observations: 124

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.604	0.604	46.358	0.000
		2	0.402	0.059	67.064	0.000
		3	0.301	0.105	81.199	0.000
		4	0.481	0.371	111.34	0.000
		5	0.224	0.388	117.95	0.000
		6	0.178	0.167	122.15	0.000
		7	0.100	-0.120	123.57	0.000
		8	0.270	0.239	134.05	0.000
		9	0.033	-0.374	134.20	0.000
		10	0.032	0.089	134.34	0.000
		11	0.090	0.048	135.46	0.000
		12	0.060	-0.038	135.97	0.000
		13	-0.100	0.033	137.63	0.000
		14	-0.114	-0.127	139.48	0.000
		15	0.201	0.017	145.25	0.000
		16	0.070	0.101	146.07	0.000
		17	-0.215	-0.017	152.02	0.000
		18	-0.222	-0.079	160.10	0.000
		19	-0.315	-0.116	174.59	0.000
		20	0.178	0.070	179.64	0.000
		21	0.297	0.206	193.00	0.000
		22	-0.200	0.126	203.30	0.000
		23	-0.345	-0.218	221.76	0.000
		24	-0.224	0.063	229.53	0.000
		25	0.316	0.113	245.36	0.000
		26	0.240	0.000	254.54	0.000
		27	-0.257	0.105	266.17	0.000
		28	-0.117	-0.323	267.30	0.000
		29	-0.170	0.457	272.64	0.000
		30	0.104	0.764	274.45	0.000

En el Cuadro 34 se observa que las correlaciones (AC) y correlaciones parciales (PAC) están fuera de las bandas de confianza, por lo que existe autocorrelación entre los errores. Asimismo, se observa que los primeros valores de AC y PAC son grandes; por su parte los valores que asume el estadístico Q son muy altos mientras que sus probabilidades asociadas son menores a 0.05. Todo lo anterior denota que el modelo presenta autocorrelación.

Métodos formales para la detección de la Autocorrelación.

Prueba Durbin–Watson.

Primero estimaremos la ecuación como ya se ha realizado anteriormente. En la barra de comandos se escribe la regresión LS X C PIB FBKF INPC TC CG, y se obtiene el siguiente Cuadro:

Cuadro 36. Resultados de la Regresión Original.

Dependent Variable: M
Method: Least Squares
Autocorrelación
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-61173.23	2205.992	-27.73048	0.0000
CP	1.98E-05	1.35E-06	14.69184	0.0000
FBKF	1.14E-05	3.30E-06	3.454748	0.0008
CG	-8.87E-07	3.47E-06	-0.255585	0.7987

R-squared	0.976706	Mean dependent var	35639.22
Adjusted R-squared	0.976123	S.D. dependent var	26300.26
S.E. of regression	4076.775	Akaike info criterion	19.49573
Sum squared resid	1.99E+09	Schwarz criterion	19.58670
Log likelihood	-1204.735	F-statistic	1677.146
Durbin-Watson stat	0.738435	Prob(F-statistic)	0.000000

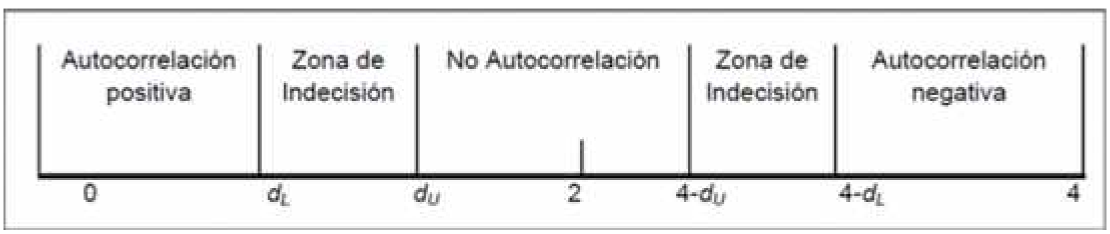
Del Cuadro 35 se desprende la siguiente ecuación de regresión:

$\hat{M}=-61173.23403 + 0.00001984597611*CP + 0.00001141058258*FBKF - 0.0000008869354734*CG$

La pantalla de regresión anterior muestra que el estadístico Durbin-Watson asciende a 0.738435. A continuación se obtienen los valores de los límites inferior (d_L) y superior (d_U), tomando en cuenta que n= 124 y k=4, se buscan los valores en la tabla *d* de Durbin Watson los cuales son 1.679 y 1.788.

Entonces se tiene que $d = 0.738435 < d_L = 1.679$, por lo que se rechaza la hipótesis nula y concluimos que existe autocorrelación positiva en el modelo.

Cuadro 37. Zonas de aceptación Durbin-Watson.



Prueba Breusch-Godfrey.

Esta prueba permite construir una hipótesis alternativa buscando el orden p (coeficiente de autocorrelación) en un proceso autoregresivo o el orden de un proceso de medias móviles (MA).

Para realizar la prueba Breusch-Godfrey en EViews 5 se estima la ecuación de regresión (ver Cuadro 27); en la ventana de la regresión vamos a: View/Residuals Test/ Serial Correlation LM Test/ Lags to include (incluir 3 ó 4 rezagos debido a que la prueba Durbin-Watson nos permite probar autocorrelación de orden 1, el correlograma de orden 2, además de que los datos son trimestrales, por lo que será preciso probar Autocorrelación de orden 3 y 4)

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación.

Ha: Autocorrelación.

La regla de decisión es la siguiente:

Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Cuadro 38. Prueba Breusch-Godfrey con 3 rezagos.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	26.67734	Probability	0.000000
Obs*R-squared	50.36729	Probability	0.000000

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Bresuch-Godfrey
Presample missing value lagged residuals set to zero.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2353.569	1867.464	1.260297	0.2101
CP	4.79E-07	1.19E-06	0.398456	0.6910
FBKF	-2.23E-07	2.72E-06	-0.081990	0.9348
CG	4.96E-06	3.68E-06	1.348359	0.1801
RESID(-1)	0.602722	0.097422	6.186697	0.0000
RESID(-2)	-0.019592	0.134060	-0.146012	0.8843
RESID(-3)	0.144393	0.126639	1.140188	0.2565

R-squared	0.406100	Mean dependent var	1.47E-11
Adjusted R-squared	0.375736	S.D. dependent var	4026.752
S.E. of regression	3181.552	Akaike info criterion	19.02292
Sum squared resid	1.18E+09	Schwarz criterion	19.18213
Log likelihood	-1172.421	F-statistic	13.33007
Durbin-Watson stat	1.892228	Prob(F-statistic)	0.000000

A partir del Cuadro 38 se puede concluir que al ser la probabilidad asociada al estadístico “F” igual a 0.0000 y por tanto menor a 0.05, se acepta la Ha y concluimos que el modelo presenta autocorrelación. Dicha conclusión se corrobora con el hecho de que la variable “RESID(-1)” es estadísticamente significativa y por tanto explica a la variable dependiente (RESID); el estadístico F de la regresión auxiliar (Prueba BG) indica que en su conjunto todas las variables independientes de la regresión auxiliar explican a la variable dependiente, en este caso, a los residuales; y el valor del R^2 no es pequeño ya que este asciende a 40.61%. Todo lo anterior corrobora que el modelo presenta autocorrelación.

Prueba ARCH LM.

La prueba de Arch LM analiza si la varianza tiene un proceso de comportamiento Arch, es decir, si depende ó no de ella misma rezagada (n) veces. Además en un proceso heterocedástico pueda presentarse la varianza del tipo Arch(n).

Para ver la prueba en EViews 5 de Arch LM Test, se estima la ecuación de regresión (ver Cuadro 27); en la ventana de la regresión vamos a: View/Residuals Test/ ARCH LM Test/ Lags to include (incluir 3 ó 4 rezagos debido a que la prueba Durbin-Watson prueba autocorrelación de orden 1, el correlograma de orden 2, además que los datos son trimestrales, por lo que será preciso probar Autocorrelación de orden 3 y 4)

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación de orden n.

Ha: Autocorrelación de orden n.

La regla de decisión es la siguiente:

Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Cuadro 39. Prueba ARCH LM CON 3 REZAGOS.

ARCH Test:

F-statistic	5.336242	Probability	0.001762
Obs*R-squared	14.56097	Probability	0.002233

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
ARCH LM Test
Sample (adjusted): 1980Q4 2010Q4
Included observations: 121 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8735015.	2609400.	3.347559	0.0011
RESID^2(-1)	0.146550	0.093412	1.568847	0.1194
RESID^2(-2)	0.247198	0.094611	2.612797	0.0102
RESID^2(-3)	0.093117	0.101076	0.921258	0.3588

R-squared	0.120339	Mean dependent var	16384073
Adjusted R-squared	0.097783	S.D. dependent var	19439015
S.E. of regression	18464167	Akaike info criterion	36.33306
Sum squared resid	3.99E+16	Schwarz criterion	36.42548
Log likelihood	-2194.150	F-statistic	5.336242
Durbin-Watson stat	2.032197	Prob(F-statistic)	0.001762

Se puede observar en el Cuadro 39 que la probabilidad asociada al estadístico F es igual a 0.011163, que es menor a 0.05, por lo tanto se acepta la Ha y concluimos que el modelo presenta autocorrelación. Por tanto, la varianza presenta un proceso de comportamiento ARCH(n), depende de ella rezagada (n) veces. Existe la posibilidad de que en el proceso heterocedástico pueda presentarse una varianza de tipo ARCH (n). Asimismo observamos que la variable "RESID^2(-2)" es estadísticamente significativa, lo cual indica que ésta explica a la variable dependiente, en este caso “RESID^2”. En caso de que no existiera autocorrelación los residuos rezagados no deberían tener relación alguna con los residuos, es decir, con la variable dependiente. El estadístico F de la regresión auxiliar (prueba ARCH LM) es estadísticamente significativo, por tanto, denota que las variables independientes explican a la variable dependiente, en caso de que no existiera autocorrelación, el estadístico F deberá ser no significativo. Por lo que respecta al valor del R^2, éste asciende a 12.03%, por lo que las variables independientes explican el 12.03% de los cambios que ocurren en la variable dependiente, en caso de que no existiera autocorrelación el R^2 deberá ser cercano o tendiente a cero. Por tanto, en este caso, a partir de esta prueba, concluimos que el modelo presenta autocorrelación.

Propuesta de corrección.

Dado que el modelo originalmente planteado presenta problemas de autocorrelación es necesario reespecificarlo. Para ello se plantea una nueva regresión que solucione el problema de la autocorrelación. La nueva regresión utiliza una forma funcional diferente (logaritmos), rezagos y un cambio importante es el ajuste del periodo, este fue delimitado a un periodo con menores cambios. También, se eliminaron y cambiaron algunas variables. El periodo que se determinó en la nueva regresión es del primer trimestre de 1996 hasta el tercer trimestre de 2007

Así, la regresión reespecificada es la siguiente:

$$LM = f\left(LX(-2) + LFBKF(-1)\right)$$

Donde:

LM= logaritmo de la M.

LX= logaritmo de las X rezagada 2 periodos.

LFBKF= logaritmo de la FBKF rezagado 1 periodo.

Estimación de la regresión.

Como se reespecificó la regresión se volverá a hacer la estimación en EViews 5. Así, en la barra de comandos escribiremos la regresión: LS LM C LX(-2) LFBKF(-1), con lo cual se desplegará una ventana como la que se muestra a continuación:

Cuadro 40. Resultados de la regresión reespecificada.

Dependent Variable: LM
Method: Least Squares
Autocorrelación: correción
Sample: 1996Q1 2007Q3
Included observations: 47

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.806518	1.511083	-3.180842	0.0027
LX(-2)	0.703391	0.066719	10.54257	0.0000
LFBKF(-1)	0.384159	0.103070	3.727156	0.0005

R-squared	0.977744	Mean dependent var	10.82022
Adjusted R-squared	0.976732	S.D. dependent var	0.293273
S.E. of regression	0.044735	Akaike info criterion	-3.314413
Sum squared resid	0.088054	Schwarz criterion	-3.196319
Log likelihood	80.88871	F-statistic	966.4944
Durbin-Watson stat	1.458460	Prob(F-statistic)	0.000000

Del Cuadro 40 se desprende la siguiente ecuación de regresión:

LM= -4.806518+ 0.703391*LX (-2)+0.384159*LFBKF(-1)

Detección de autocorrelación.

A continuación se procederá a determinar si con la reespecificación del modelo se corrigió el problema de la Autocorrelación.

1. **Prueba Breusch-Godfrey.**

Para realizar la prueba en EViews 5 de Breusch-Godfrey, se estima la ecuación de regresión (ver Cuadro 31); en la ventana de la regresión vamos a: View/Residuals Test/ Serial Correlation LM Test/ Lags to include (3 ó 4)

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación.

Ha: Autocorrelación.

La regla de decisión es la siguiente:

Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Cuadro 41. Corrección prueba Breusch-Godfrey con 3 rezagos.

Breusch-Godfrey Serial Correlation LM Test:

F-statistic	1.917745	Probability	0.141779
Obs*R-squared	5.783500	Probability	0.122527

Test Equation:
Dependent Variable: RESID
Method: Least Squares
Breusch-Godfrey
Presample missing value lagged residuals set to zero

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.516374	1.620362	0.936195	0.3547
LX(-2)	0.067377	0.071282	0.953540	0.3459
LFBKF(-1)	-0.106519	0.110881	-0.961565	0.3419
RESID(-1)	0.380503	0.172323	2.208089	0.0329
RESID(-2)	-0.234497	0.166083	-1.411873	0.1655
RESID(-3)	0.009227	0.160810	0.057377	0.9545

R-squared	0.123055	Mean dependent var	3.78E-16
Adjusted R-squared	0.016111	S.D. dependent var	0.043752
S.E. of regression	0.043398	Akaike info criterion	-3.318065
Sum squared resid	0.077219	Schwarz criterion	-3.081576
Log likelihood	53.97453	F-statistic	1.150647
Durbin-Watson stat	1.891501	Prob(F-statistic)	0.349598

En el Cuadro 41 se observa que la probabilidad asociada al estadístico F es 0.141779 lo cual es mayor a 0.05, por lo que se acepta la Ho y se concluye que no existe autocorrelación. Asimismo, el estadístico F de la regresión auxiliar es no significativo, indicando que las variables independientes en su conjunto no explican a la variable dependiente; el valor del R^2 es bajo. Por tanto, mediante esta prueba se concluye que el modelo no presenta autocorrelación.

2. **Prueba ARCH LM.**

Para realizar la prueba en EViews 5 de Arch LM Test, se estima la ecuación de regresión; en la ventana de la regresión vamos a: View/Residuals Test/ ARCH LM Test/ Lags to include (3 ó 4)

La prueba de hipótesis es la siguiente:

Ho: No Autocorrelación de orden n.

Ha: Autocorrelación de orden n.

La regla de decisión es la siguiente:

Si la probabilidad asociada al estadístico F ≥ 0.05, no se rechaza la Ho.

Si probabilidad asociada al estadístico F <0.05, se rechaza la Ho.

Cuadro 42. Prueba ARCH LM.

ARCH Test:			
F-statistic	0.346474	Probability	0.791861
Obs*R-squared	1.114406	Probability	0.773598

Test Equation:
Dependent Variable: RESID^2
Method: Least Squares
ARCH LM Test
Sample (adjusted): 1996Q4 2007Q3
Included observations: 44 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.001785	0.000578	3.086756	0.0037
RESID^2(-1)	0.156747	0.160375	0.969722	0.3203
RESID^2(-2)	0.023638	0.161128	0.146704	0.8841
RESID^2(-3)	-0.034636	0.157558	-0.219832	0.8271

R-squared	0.025327	Mean dependent var	0.001975
Adjusted R-squared	-0.047773	S.D. dependent var	0.002238
S.E. of regression	0.002291	Akaike info criterion	-9.233391
Sum squared resid	0.000210	Schwarz criterion	-9.071192
Log likelihood	207.1346	F-statistic	0.346474
Durbin-Watson stat	1.971178	Prob(F-statistic)	0.791861

En el Cuadro 42 observamos que la probabilidad asociada al estadístico F es de 0.791861, como es mayor a 0.05, se acepta la Ho y concluimos que no existe autocorrelación en el modelo. Con ello concluimos que la varianza no presenta un proceso de comportamiento ARCH(n), es decir, no depende de ella misma rezagada en el tiempo, se descarta la posibilidad de que exista un proceso heteroscedástico que pueda hacer que la varianza sea del tipo ARCH(n). La no autocorrelación se corrobora en el hecho de que ninguna de las variables independientes incluidas en la regresión auxiliar es estadísticamente significativa; el estadístico F de la regresión auxiliar indica que las variables independientes en su conjunto no explican a la dependiente; y el valor del R^2 es insignificante, a saber, 2.53%.

Correlograma.

Para realizar la prueba en EViews 5 de Correlogram Q Statistics se estima la ecuación de regresión; en la ventana de la regresión vamos a: View/Residuals Test/ Correlogram Q Statistics / Lags to include 30.

Cuadro 43. Correlograma.

Correlograma Q
Sample: 1995Q1 2007Q3
Included observations: 47

	AC	PAC	Q-Stat	Prob
1	0.253	0.253	3.2015	0.074
2	-0.118	-0.194	3.9141	0.141
3	0.090	0.006	4.3427	0.227
4	[illegible]	0.268	[illegible]	0.121
5	0.207	0.441	9.6540	0.086
6	-0.256	0.028	13.326	0.038
7	0.009	0.107	13.331	0.064
8	0.278	0.053	17.381	0.022
9	-0.127	-0.169	18.259	0.026
10	-0.346	-0.272	26.297	0.003
11	-0.118	0.021	27.103	0.004
12	0.194	0.082	29.560	0.003
13	-0.106	-0.148	30.425	0.004
14	0.140	0.095	31.796	0.004
15	-0.005	-0.214	31.800	0.007
16	0.205	0.033	34.931	0.004
17	-0.134	-0.059	36.305	0.004
18	-0.134	-0.030	37.727	0.004
19	0.010	0.028	37.736	0.006
20	0.199	-0.095	[illegible]	0.004
21	0.079	0.032	41.668	0.005
22	-0.110	-0.055	42.780	0.005
23	0.014	0.050	42.796	0.007
24	0.132	0.001	44.532	0.007
25	-0.011	0.045	44.546	0.009
26	-0.054	-0.032	44.866	0.012
27	-0.061	-0.197	45.296	0.015
28	0.008	0.106	45.302	0.021
29	0.003	0.160	45.302	0.027
30	0.000	0.047	45.302	0.036

En el Cuadro 43 se observa que tanto en Autocorrelación (AC) como en Autocorrelación Parcial (PAC) las barras se salen de las bandas de confianza, asimismo, los primeros valores de AC y PAC no son pequeños; los valores del estadístico Q son altos y en la mayoría de los casos sus probabilidades asociadas son menores a 0.05. Por lo que, a partir del correlograma se concluye que la autocorrelación en el modelo aun persiste.

Sin embargo, si bien el correlograma nos indica que aún persiste la autocorrelación en el modelo, las pruebas Breusch-Godfrey (BG) y ARCH LM nos indican lo contrario. Para fines prácticos nos quedaremos con los resultados de las pruebas BG y ARCH LM ya que, por construcción, son pruebas más robustas y por ende, más confiables. En este sentido, concluimos que las operaciones realizadas fueron idóneas para resolver el problema de la autocorrelación.

En resumen, con estos tres ejercicios ahora el lector dispone de más métodos tanto para identificar como para eliminar o reducir la autocorrelación. Así, estos métodos complementan los descritos en la exposición teórica del tema.

VI.2.6.- REACTIVOS PARA REAFIRMAR LOS CONOCIMIENTOS.

A.- Examen de autoevaluación

Tema: Efecto sobre los estimadores de la violación de los supuestos del modelo de regresión lineal: Mínimos cuadrados ordinarios, Autocorrelación.

A.-Conteste con una “X” en **SI** cuando la frase afirmativa sea verdadera y también con una “X” en NO cuando la frase afirmación sea falsa:

1.-Uno de los supuestos es que las Ui si están correlacionadas: Si___; No____.

2.- La hipótesis nula es que los Ui son dependientes, es decir, r mayor que cero : SI_____;NO_____.

3.- Una autocorrelación refleja que el valor del año anterior de la variable influye en su valor de este año: Si_____; No________

4.-Cuando la d= 2 y r= 0 se dice que no hay autocorrelación: SI:_______; NO:___________

5.- En opinión de Dominick Salvatore:*la autocorrelación se corrige estimando **ρ** por ser el indicador de la autocorrelación serial:* SI:__; NO__.

6.- El cálculo de *ρ(coeficiente de autocorrelación)* y de r *(coeficiente de correlación)* es el mismo*:* SI:__; NO__.

7.- *ρ y r, ambas expresan relaciones pero de variables distintas:* SI:__; NO__.

8.-La prueba d de Durbin-Watson no se usa para identificar la existencia de autocorrelación: SI:__; NO__.

9.- La prueba h de Durbin no se usa para identificar la existencia de autocorrelación cuando el modelo no incluye el término constante: SI:__; NO__.

10.- La prueba h de Durbin es recomendable aplicar cuando la variable

Endógena se usa como exógena en la ecuación de regresión. SI:__; NO__.

11.- La prueba h de Durbin se usa cuando la variable endógena ya es estocástica: SI:__; NO__.

12.- Las variables instrumentales se usan para corregir la autocorrelación cuando ésta se identifica con h que existe: SI:__; NO__.

13.-Al ser h N(0,1) su valor se compara con la Z de la distribución normal, dado cierto valor del nivel de significación para ver si existe autocorrelación: SI:__; NO__.

14.- Los modelos con retardos distribuidos finitos sirven para determinar hasta dónde llega el efecto de la variable independiente sobre la dependiente en el tiempo: SI:__; NO__.

15.- El contraste de exogeneidad de **HAUSMAN** se usa para detectar la existencia de variables en un modelo estocásticas: SI__; NO__.

16.- El esquema de **KOYCK, dentro de los Modelos de retardos infinitos** es el que más se utiliza, cuando existe la conjetura de que los parámetros decaen geométricamente, en otras palabras, el efecto de la variables exógena se reduce exponencialmente conforme pasa el tiempo: SI: __; NO__.

VI.3.- MULTICOLINEALIDAD.

Se dice que existe multicolinealidad cuando dos o más variables explicativas están **altamente** correlacionadas en el modelo de regresión múltiple; esta alta correlación impide conocer el efecto individual de cada una de estas variables explicativas sobre la variable dependiente.

VI.3.1.- Consecuencias de la correlación entre variables explicativas.

Los coeficientes estimados con el método de mínimos cuadrados ordinarios, en opinión de D. Salvatore (idem,151), "pueden ser estadísticamente insignificantes", aun cuando se vea que R^2 tenga valores muy altos y, lo que es más importante, los coeficientes estimados aun siguen siendo INSESGADOS. Es más, Salvatore menciona que si el propósito principal de la regresión es el **PRONÓSTICO** "la multicolinealidad no es un problema si el mismo patrón de multicolinealidad persiste durante el período pronosticado".

VI.3.2.- ¿Cómo se identifica la multicolinealidad?

Aquí antes de iniciar la explicación de los métodos que podemos usar, he juzgado conveniente traer a colación **lo señalado por** Gujarati (1990:258) sobre la heteroscedasticidad en el sentido de que no es tan fácil detectarla, cuando dice que " **no existen reglas fijas y seguras para detectarla, sino sólo unas cuantas reglas generales". ¿**En el caso de la multicolinealidad debemos interpretar igualmente a Wooldridge **(2009: xiv) ?** quien dice que " mi opinión – de que la multicolinealidad es aún un tema poco comprendido y de que las afirmaciones que dicen que uno puede detectar y corregir multicolialidad están equivocadas-no ha cambiado".

Con esas referencias y actuando con las reservas del caso, podemos decir que existe multicolilealidad entre las variables explicativas:

1. Cuando se observa que *alguno* o ninguno de los coeficientes de las variables explicativas *es estadísticamente significativo*, además de que R^2 resulta alto y F muestra que en conjunto si son significativos estadísticamente. Carrascal (2001:162). En resumen, *cuando la significación estadística de las t´s con F es distinta y cuando R^2 es alto*.

2. También se detecta la multicolinealidad cuando se obtienen elevados *coeficientes de correlación simple o parcial*, entre las variables explicatorias; sin embargo, esto no es muy seguro porque puede presentarse multicolinealidad "suficiente aun si los coeficientes de correlaciones simples o parciales son relativamente bajos (menores que 0.5)". Derivado de lo anterior es que Carrascal (2001:174) propone calcular la *matriz de correlaciones* **entre cada par de regresores**, es decir hacer análisis de correlación simple; si la correlación es elevada (próxima a $\pm$ 1) es indicativo de que hay multicolinealidad.

3.- Con la prueba de **Theil**, estableciendo que $0 \leq Theil \leq 1$, de manera que cuando:

Ho: Theil=0, se dice que no existe multicolinealidad entre las variables exógenas;

Ha: Theil=1, se dice que si existe multicolinealidad entre ellas

Su fórmula es Theil= R^2-∑(R^2-R^2_i) ; donde: R^2 = Coeficiente de determinación múltiple, R^2_{ij}=coeficiente de determinación de la variable exógena i-ésima con la variable endógena j-ésima.

Regla de decisión:

Cuando Theil ≤ 0.5, no hay multicolinealidad; y
Cuando Theil ≥ 0.5, si hay multicolinealidad

Ejemplo: Sea el caso en que existen la variable endógena (1) explicada por dos exógenas (2 y 3), tal que

R^2_{123}= 0.870586; $R^2_{12=}$ 0.681399; R^2_{13}= 0.816927, entonces:

Theil: 0.870586- ((0.870586-0.681399)+(0.870586-0.816927))=0.62774, se dice que existe multicolinealidad entre las dos variables exógenas (2 y 3).

4.- Con el **Factor de Inflación de Varianza,**(FIV), cuya fórmula es FIV=1/1-R^2 y estableciendo que:

Ho: Cuando FIV es menor que 10, no hay multicolinealidad entre las variables exógenas; y

Ha: Cuando FIV es mayor que 10, si hay multicolinealidad entre las variables exógenas.

Ejemplo: Si del ejemplo anterior tomamos R^2=0.870586, entonces
FI= 1/1-0.870586=7.7271, se dice que no existe multicolinealidad entre ellas.

VI.3.3.-Métodos para disminuirla o eliminar la multicolinealidad.

a) Se amplía el tamaño de los datos muestrales;
b) Utilizar información a priori;
c) Se *transforma* la relación funcional: incrementando o deflactando las variables del modelo.
d) Se omite una de las variables altamente colineales. En este caso puede surgir un problema de especificación o error si la teoría señala que dicha variable omitida se debe incluir en el modelo, por ello no es recomendable.

NOTA: La *transformación* de variables incluidas en el modelo para que la nuevas variables transformadas presenten correlaciones lineales más bajas se hace incrementando las variables, como ya se indicó y, en el caso de la deflactación de las mismas, se hace con INPC u otro apropiado, de modo que los datos del modelo ahora se expresan en valores reales y no nominales y con ello se elimina la multicolinealidad.

VI.3.4.-APLICACIÓN CON EVIEWS

A.-Ejemplos adicionales utilizando Eviews 5.

En beneficio del mayor y mejor aprendizaje del lector de esta metodología, es conveniente que compare los métodos expuestos en la sección teórica de este tema, con los que a continuación se expondrán; todo ello con el objetivo de que incremente su acervo, evalué cada uno de ellos y se forme un criterio sólido sobre lo robusto de cada uno de los métodos, así como cuándo aplicar cada uno de ellos.

La ilustración de presencia y corrección de multicolinealidad se realizará mediante el uso de EViews 5 utilizando datos actuales de la economía mexicana.

Ejemplo 1: Consumo Gubernamental.

Basándonos en la teoría Keynesiana se realizará una regresión donde el Consumo Gubernamental (CG) esté explicado por el Producto Interno Bruto (PIB), la Formación Bruta de Capital Fijo (FBKF) y las Exportaciones (X):

$$\hat{C}G = \hat{a} + \hat{b}_1 PIB + \hat{b}_2 FBKF + \hat{b}_3 X + \hat{e}$$

La prueba de hipótesis para el caso de multicolinealidad será la siguiente:

Ho: No multicolinealidad.

Ha: Multicolinealidad.

Pruebas para detectar multicolinealidad.

1. **Coeficiente de determinación alto y T´s no significativas.**

En presencia de multicolinealidad es común encontrar que uno o más coeficientes parciales son de manera individual estadísticamente no significativos con base en la prueba *t*. Mientras que el R^2 en esta situación es alto y la prueba F resulta ser estadísticamente significativa.

A continuación se realizará la regresión en EViews para analizar los resultados.

Para la obtención del Cuadro 2 se realizó la regresión escribiendo en la barra de comandos LS CG C PIB FBKF X, tal como se muestra en el Cuadro 1.

Cuadro 1. Comando de regresión.

Cuadro 2. Resultados de la regresión.

Dependent Variable: CG
Method: Least Squares
Regresión Original
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.93E+08	18877883	36.68549	0.0000
PIB	8583.449	4098.835	2.095339	0.0382
FBKF	-3.065359	0.903870	-3.391370	0.0009
X	-5230.579	4171.411	-1.253911	0.2123

R-squared	0.515574	Mean dependent var	8.11E+08
Adjusted R-squared	0.503463	S.D. dependent var	1.50E+08
S.E. of regression	1.06E+08	Akaike info criterion	39.82553
Sum squared resid	1.35E+18	Schwarz criterion	39.91650
Log likelihood	-2465.183	F-statistic	42.57191
Durbin-Watson stat	3.305633	Prob(F-statistic)	0.000000

Se puede ver en el Cuadro 2 que el R-squared (R^2) asume un valor de 0.5155, este es un valor relativamente considerable, por lo que concluimos que las variables independientes están explicando al Consumo de Gobierno (CG) en un 51.55%.

Ahora bien, como se recordará, para determinar si los coeficientes estimados son estadísticamente significativos, es decir, que su valor sea distinto de cero, nos basamos en sus probabilidades asociadas. Para ello se establecen los siguientes criterios:

- Si la probabilidad asociada al coeficiente es menor o igual a 0.05 se dice que éste es estadísticamente significativo, por lo que, la variable independiente explica a la dependiente.

- Si la probabilidad asociada al coeficiente es mayor a 0.05 se dice que éste es estadísticamente no significativo, por lo que, la variable independiente no explica a la dependiente.

Por lo que respecta al estadístico F, planteamos los siguientes criterios de decisión:

- Si la probabilidad asociada al estadístico F es menor o igual a 0.05, se concluye que en su conjunto todas las variables independientes están explicando a la variable dependiente.

- Si la probabilidad asociada al estadístico F es mayor a 0.05, se concluye que en su conjunto todas las variables independientes no están explicando a la variable dependiente.

En nuestro caso, observamos (ver Cuadro 2) que el R^2 es relativamente considerable, mientras que el estadístico F nos indica que en su conjunto todas las variables independientes explican a la variable dependiente. Sin embargo, las variables PIB y FBKF son estadísticamente significativas ya que su probabilidad es menor a 0.05; mientras que la probabilidad asociada a la variable X es mayor a 0.05, indicando que ésta no explica a la variable dependiente, es decir, es no estadísticamente significativa. Estos resultados indican la posibilidad de alta multicolinealidad.

2. Matriz de correlaciones.

Para determinar si existe alta multicolinealidad por medio de la matriz de correlaciones se sugiere observar el coeficiente de correlación entre dos regresoras, si alguno de estos es mayor al coeficiente de determinación de la regresión original (R^2), entonces tenemos presencia de alta multicolinealidad.

Planteamos la siguiente prueba de hipótesis.

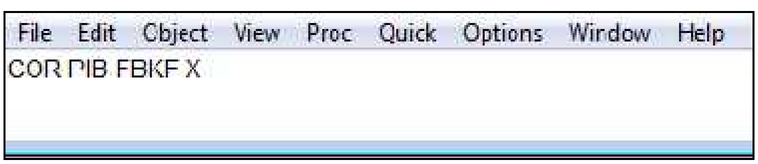
Ho: No multicolinealidad.
Ha: Multicolinealidad.

Regla de decisión:

Se acepta Ho si $R^2 > r_{x1}, r_{x2}, \ldots, r_{xn}$
Se acepta Ha si $R^2 < r_{x1}, r_{x2}, \ldots, r_{xn}$

Para obtener la matriz de correlaciones en Eviews 5, en la barra de comandos se escribe **cor PIB FBKF X**, como se muestra en el Cuadro 3, que finalmente proporcionará los resultados del Cuadro 4.

Cuadro 3. Comando matriz de correlaciones.



Cuadro 4. Matriz de correlaciones.

	PIB	FBKF	X
PIB	1.000000	-0.337159	0.996041
FBKF	-0.337159	1.000000	-0.348175
X	0.996041	-0.348175	1.000000

En el Cuadro 4 observamos que el coeficiente de correlación de las variables PIB vs X es 0.9960 > 0.51 del R^2 que se obtuvo del Cuadro 2, entonces se concluye que se está ante presencia de multicolinealidad.

3. Regresiones auxiliares (Prueba de Klein).

Se realizarán tres regresiones auxiliares, en donde cada variable independiente tomará el papel, ahora, de variable dependiente, con la finalidad de comparar los R^2 de cada una de ellas respecto al R^2 de la regresión original.

La regla de decisión será la siguiente:

- Si al menos un R^2 de una regresión auxiliar es mayor al R^2 de la regresión original, se dice que existe alta multicolinealidad.
- Si en ninguna de las regresiones auxiliares el R^2 es mayor al R^2 de la regresión original, se dice que no existe alta multicolinealidad.

Así, mediante los siguientes comandos de Eviews 5 se formulan las regresiones auxiliares:

Regresión Auxiliar 1: **LS PIB C FBKF X**

Regresión Auxiliar 2: **LS FBKF C PIB X**

Regresión Auxiliar 3: **LS X C PIB FBKF**

En los Cuadros 5, 6 y 7 se muestra los resultados de las regresiones auxiliares.

Cuadro 5. Resultados de la Regresión auxiliar 1.

Dependent Variable: PIB
Method: Least Squares
Regresión Auxiliar 1
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coeficient	Std. Error	t-Statistic	Prob.
C	1701.280	389.0851	4.372513	0.0000
FBKF	2.55E-05	1.99E-05	1.280844	0.2027
X	1.013221	0.008677	116.7710	0.0000

R-squared	0.992204	Mean dependent var	35639.22
Adjusted R-squared	0.992075	S.D. dependent var	26383.26
S.E. of regression	2348.679	Akaike info criterion	18.38499
Sum squared resid	6.67E+08	Schwarz criterion	18.45322
Log likelihood	-1136.869	F-statistic	7699.909
Durbin-Watson stat	0.381662	Prob(F-statistic)	0.000000

Cuadro 6. Resultados de la Regresión auxiliar 2.

Dependent Variable: FBKF
Method: Least Squares
Regresión Auxiliar 2
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8125510.	1749108.	4.645516	0.0000
PID	524.4053	409.4042	1.280044	0.2027
X	-681.1031	414.9561	-1.641398	0.1033

R-squared	0.132981	Mean dependent var	4073814.
Adjusted R-squared	0.118650	S.D. dependent var	11344980
S.E. of regression	10650694	Akaike info criterion	35.22404
Sum squared resid	1.37E+16	Schwarz criterion	35.29228
Log likelihood	-2180.091	F-statistic	9.279320
Durbin-Watson stat	0.045216	Prob(F-statistic)	0.000178

Cuadro 7. Resultados de la Regresión auxiliar 3.

Dependent Variable: X
Method: Least Squares
Regresión Auxiliar 3
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1341.965	392.9036	-3.415463	0.0009
PIB	0.973271	0.008373	116.7710	0.0000
FBKF	-3.20E-06	1.95E-05	-1.641398	0.1033

R-squared	0.992270	Mean dependent var	33392.56
Adjusted R-squared	0.992143	S.D. dependent var	26035.35
S.E. of regression	2307.815	Akaike info criterion	18.34989
Sum squared resid	6.44E+08	Schwarz criterion	18.41812
Log likelihood	-1134.693	F-statistic	7766.580
Durbin-Watson stat	0.379709	Prob(F-statistic)	0.000000

Así, en base a los resultados de las regresiones auxiliares, se puede observar que tanto en la regresión auxiliar 1 como en la 3, el R^2 es mayor al R^2 de la regresión original, concluyéndose que existe alta multicolinealidad.

4. **Prueba de Theil.**

La prueba Theil es la diferencia entre la variabilidad explicada por el modelo completo y la suma de cada una de las aportaciones de los regresores al modelo. Para ello se utilizarán los R^2 obtenidos en las regresiones auxiliares anteriores

Para evaluar los resultados mediante esta prueba es necesario formular la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: Multicolinealidad.

La regla de decisión es:

- Si Theil ≤ a 0.5 se acepta la Ho.
- Si Theil > a 0.5 no se acepta la Ho.

La fórmula para determinar la prueba es la siguiente:

Theil= $R^2 - \sum(R^2 - R^2_i)$.

Donde:

R^2: coeficiente de determinación de la regresión original.
$R^2_{i:}$ coeficiente de determinación de las regresiones auxiliares.

Se resta el R^2 del modelo original de cada una de las regresiones auxiliares.

R^2 original = 0.515574

R^2 regresión auxiliar 1= 0.992204

R^2 regresión auxiliar 2= 0.132981

R^2 regresión auxiliar 3= 0.992270

Theil= (0.515574 – ∑ ((0.515574 - 0.992204) + (0.515574 - 0.132981) + (0.515574 - 0.992270))) = **0.898101.**

Tomando el resultado en valor absoluto, tenemos que como Theil es igual a 0.898101, es decir, mayor a 0.5, entonces no aceptamos la hipótesis nula de no multicolinealidad.

5. Factor de Inflación de Varianza (FIV).

Basado en las regresiones auxiliares, el FIV, muestra como la varianza de un estimador se infla por la presencia de multicolinealidad. Si lo que se quiere es determinar de manera conjunta si el modelo presenta multicolinealidad, entonces, se calcula el FIV para la regresión original. Asimismo, si buscamos determinar que variables es causante de multicolinealidad, entonces se utilizan las regresiones auxiliares, para el cálculo del FIV.

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: multicolinealidad.

La Regla de decisión es:

- Si el valor de FIV es ≤ a 10 se acepta Ho.
- Si el valor de FIV es > a 10 no se acepta Ho.

La fórmula es:

$$FIV = \frac{1}{(1 - R^2)}$$

Determinación de multicolinealidad de manera conjunta:

$$FIV = \frac{1}{(1 - 0.515574)} = 2.06$$

Así, como el valor del FIV es de 2.06, entonces aceptamos Ho. Por tanto, se concluye que, de acuerdo a esta prueba, no existe alta multicolinealidad en el modelo.

Ahora estimaremos el FIV para cada una de las variables independientes para determinar cuál o cuáles son las variables independientes causantes de la alta multicolinealidad. Para ello se tomarán los R^2 correspondientes de las ecuaciones auxiliares previamente realizadas.

R^2 regresión auxiliar 1= 0.992204

R^2 regresión auxiliar 2= 0.132981

R^2 regresión auxiliar 3= 0.992270

A continuación se muestran los valores del FIV para cada regresión auxiliar:

$$FIVreg.auxiliar\ 1 = \frac{1}{(1 - 0.992204)}$$

$$\boldsymbol{FIV = 128.270}$$

$$FIVreg.auxiliar\ 2 = \frac{1}{(1 - 0.132981)}$$

$$\boldsymbol{FIV = 1.1527}$$

$$FIVreg.auxiliar\ 3 = \frac{1}{(1 - 0.992270)}$$

$$FIV = 129.366$$

Los factores de inflación de varianza de la regresiones auxiliares 1 y 3 son mayores a 10, por lo que no se acepta la hipótesis de no multicolinealidad; mientras que en la regresión auxiliar 2, el FIV es menor a 10, lo que se traduce en que la variable X no es la que produce la presencia de multicolinealidad.

6. Prueba de Farrar y Glouber.

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: multicolinealidad.

La regla de decisión es:

- Si el valor de X^2 es mayor o igual al valor de G se acepta la Ho.
- Si el valor de X^2 es menor al valor de G no se acepta la Ho.

La G denota el valor del estadístico construido por Farrar y Glouber.

La fórmula de esta prueba es la siguiente**:**

$$G = -\left(N - \frac{1}{6}(2K+5)\right) In \ |Rxx|$$

$$G \sim x^2$$

Grados de libertad = $n = k * \frac{k-1}{2}$

Donde:

$|Rxx|$ = Matriz de correlación de las variables independientes del valor absoluto.

N= número de observaciones.

K=Variables independientes incluyendo la constante.

Grados de libertad= $4 * \frac{4-1}{2}$= 6

X^2 con 6 grados de libertad= 12.59

$|Rxx|$=.00068

$$G = -\left[124 - \frac{1}{6}(2 * 4 + 5)\right] ln|0.0068|$$

$$G = 263.6780$$

Como X^2= 12.59 < G = 263.6780, entonces se concluye que existe multicolinealidad.

Finalmente para este ejemplo se puede decir que el grado de correlación entre las variables independientes es alto ya que se demostró mediante diversas pruebas.

Medidas de corrección.

1. Reespecificar la forma funcional, puede cambiarse por una regresión lineal-logarítmica y logarítmica-lineal. Debe de tomarse en cuenta que si la forma funcional cambia, también cambia la interpretación de los coeficientes.

2. Aumentar o disminuir el número de observaciones. Si las muestras son mayores a 30 observaciones la varianza de los estimadores se acercará más al verdadero valor.

3. Reespecificar la regresión, donde se puede incluir o eliminar variables, para así lograr que las variables expliquen mejor a la regresión.

Aplicando los criterios de corrección 1 y 3, ahora se realizará el siguiente modelo. En este caso, se eliminaron las variables Exportaciones y Formación Bruta de Capital Fijo y se incluyó la variable Índice Nacional de Precios al Consumidor (INPC) en logaritmos. Por su parte la variable dependiente se transformó a logaritmos. El nuevo modelo planteado es el siguiente:

$$LCG = \hat{a} - \hat{b} * PIB + LINPC + \hat{e}$$

Pruebas para detectar la presencia de multicolinealidad.

1. Coeficiente de determinación alto y T´s no significativas.

La regresión que se introducirá en la barra de comandos de EViews será: ***LS LCG C PIB LINPC***, para obtener los resultados del Cuadro 8. Recuérdese que para generar las variables en logaritmos, en la barra de comandos de EViews se escribe: genr lcg=log(cg), con lo cual se obtendrá el logaritmo de la variable CG, para el resto de las variables aplica el mismo procedimiento.

Cuadro 8. Resultados de la regresión.

Dependent Variable: LCG
Method: Least Squares
Regresión correción
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	20.32058	0.020104	1010.794	0.0000
PIE	2.75E-06	7.08E-07	3.881260	0.0002
LINPC	0.031606	0.008396	3.764352	0.0003

R-squared	0.510638	Mean dependent var	20.49709
Adjusted R-squared	0.502550	S.D. dependent var	0.187828
S.E. of regression	0.132476	Akaike info criterion	-1.180941
Sum squared resid	2.123523	Schwarz criterion	-1.112708
Log likelihood	76.21833	F-statistic	63.13042
Durbin-Watson stat	3.219403	Prob(F-statistic)	0.000000

Así, se observa que los coeficientes son estadísticamente significativos ya que sus probabilidades asociadas son menores a 0.05. Asimismo, la probabilidad asociada al estadístico F es menor a 0.05, indicando que las variables independientes explican en su conjunto a la variable dependiente. Por lo que respecta al R^2, su valor es relativamente considerable. Por tanto, concluimos que en general el modelo no presenta multicolinealidad.

Matriz de correlaciones.

Para obtener la matriz de correlación se llevaron a cabo los mismos pasos que se especifican en los Cuadros 3 y 4. Los resultados son los siguientes:

Cuadro 9. Matriz de correlaciones.

	PIF	LINPC
PIE	1.000000	0.768506
LINPC	0.768506	1.000000

Para determinar si existe multicolinealidad por medio de la matriz de correlaciones planteamos la siguiente prueba de hipótesis.

Ho: No multicolinealidad
Ha: Multicolinealidad

Regla de decisión:

Se acepta Ho si $R^2 > r_{x1}, r_{x2}, \ldots, r_{xn}$
Se acepta Ha si $R^2 < r_{x1}, r_{x2}, \ldots, r_{xn}$

Por tanto, como el coeficiente de correlación entre las variables independientes es mayor que el R^2, no se acepta la hipótesis nula de no multicolinealidad.

Regresiones Auxiliares (Prueba de Klein).

Las regresiones auxiliares para este caso son:

Regresión auxiliar 1**: LS PIB C LINPC**

Regresión auxiliar 2: **LS LINPC C PIB**

Los resultados se muestran a continuación:

Cuadro 10. Resultados de la Regresión Auxiliar 1

Dependent Variable: PIB
Method: Least Squares
Regresión auxiliar 1
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12954.69	2289.284	5.658842	0.0000
LINPC	9118.944	687.3692	13.25644	0.0000

R-squared	0.590602	Mean dependent var	35639.22
Adjusted R-squared	0.587246	S.D. dependent var	26383.26
S.E. of regression	16950.16	Akaike info criterion	22.32994
Sum squared resid	3.51E+10	Schwarz criterion	22.37543
Log likelihood	-1382.456	F-statistic	175.9985
Durbin-Watson stat	0.040534	Prob(F-statistic)	0.000000

Cuadro 11. Resultados de la Regresión Auxiliar 2.

Dependent Variable: LINPC
Method: Least Squares
Regresión auxiliar 2
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.179399	0.216168	0.829906	0.4082
PIB	6.48E-05	4.88E-06	13.26644	0.0000

R-squared	0.590602	Mean dependent var	2.487626
Adjusted R-squared	0.587246	S.D. dependent var	2.223470
S.E. of regression	1.428439	Akaike info criterion	3.567109
Sum squared resid	248.9507	Schwarz criterion	3.612597
Log likelihood	-219.1607	F-statistic	175.9985
Durbin-Watson stat	0.026230	Prob(F-statistic)	0.000000

La regla de decisión es la siguiente:

- Si al menos un R^2 de una regresión auxiliar es mayor al R^2 de la regresión original, se dice que existe alta multicolinealidad.

- Si en ninguna de las regresiones auxiliares el R^2 es mayor al R^2 de la regresión original, se dice que no existe alta multicolinealidad.

Así, dado que los R^2 de las dos regresiones auxiliares son mayores que el R^2 del nuevo modelo, se concluye, mediante esta prueba, que el modelo presenta alta multicolinealidad.

Prueba 4. Regla de Theil.

Para evaluar los resultados mediante esta prueba es necesario formular la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: Multicolinealidad.

La regla de decisión es:

- Si Theil ≤ a 0.5 se acepta la Ho.
- Si Theil > a 0.5 no se acepta la Ho.

La fórmula para determinar la prueba es la siguiente:

$Theil= R^2-\sum(R^2-R^2_i)$.

Donde:

R^2: coeficiente de determinación de la regresión original.
R^2_i: coeficiente de determinación de las regresiones auxiliares.

Se resta el R^2 del modelo original de cada una de las regresiones auxiliares. Así:

R^2 original = 0.510638

R^2 regresión auxiliar 1= 0.590602

R^2 regresión auxiliar 2= 0.590602

Theil =(0.510638-(∑(0.510638-0.590602)+ (0.510638-0.590602)))= **0.670566**

Tomando el resultado en valor absoluto, tenemos que Theil es igual a 0.670566, como es mayor a 0.5 aceptamos la hipótesis de que existe alta multicolinealidad.

Prueba 5. Factor de Inflación de Varianza (FIV).

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: multicolinealidad.

La Regla de decisión es:

- Si el valor de FIV es ≤ a 10 se acepta Ho.
- Si el valor de FIV es > a 10 no se acepta Ho.

La fórmula es:

$$FIV = \frac{1}{(1-R^2)}$$

Los resultados son los siguientes:

R^2 de la ecuación original = 0.510638

R^2 regresión auxiliar 1= 0.590602

R^2 regresión auxiliar 2= 0.590602

$$FIV\ regaux1 = \frac{1}{(1-0.590602)}$$

$$FIV\ regaux1 = \mathbf{2.4426}$$

$$FIV\ regaux2 = \frac{1}{(1-0.590602)}$$

$$FIV\ regaux2 = \mathbf{2.4426}$$

De acuerdo con el Factor de Inflación de Varianza, los valores que presentan las regresiones auxiliares son menores a 10, entonces se acepta la hipótesis nula de no multicolinealidad.

6. Prueba de Farrar y Glouber.

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: Multicolinealidad.

La regla de decisión es:

- Si el valor de X^2 es mayor o igual al valor de G se acepta la Ho.
- Si el valor de X^2 es menor al valor de G no se acepta la Ho.

La G denota el valor del estadístico construido por Farrar y Glouber.
La fórmula de esta prueba es la siguiente**:**

$$G = -\left(N - \frac{1}{6}(2K+5)\right) In\ |Rxx|$$

$G \sim x^2$

Grados de libertad = $n = k * \frac{k-1}{2}$

Donde:

$|Rxx|$ = Matriz de correlación de las variables independientes del valor absoluto.

N= número de observaciones.

K=Variables independientes incluyendo la constante.

Grados de libertad= $3 * \frac{3-1}{2}$= 3

X^2 con 3 grados de libertad= 7.81

$|Rxx|$=0.409398

$$G = -\left[124 - \frac{1}{6}(2*3+5)\right] ln|0.409398|$$

$$G = 47.3828$$

Entonces como $X^2 = 7.81 < G = 47.38$, se acepta la Ha, por tanto, existe alta multicolinealidad.

Como se desprende de los resultados obtenidos, algunas pruebas nos indican que la multicolinealidad disminuyó, sin embargo, algunas otras nos indican su alta persistencia. Por lo que, para fines prácticos, el estudioso o investigador que se encuentre en alguna situación similar, es decir, de contradicción entre los resultados de las pruebas, para el caso de multicolinealidad, nos quedaremos con la prueba donde las variables independientes son estadísticamente significativas al lado de una prueba F significativa y un R^2 alto.

Ejemplo 2. Exportaciones de bienes y servicios.

Se realizara una regresión en donde las exportaciones estén en función del Producto Interno Bruto (PIB), la Formación Bruta de Capital (FBKF), el Índice Nacional de Precios al Consumidor (INPC), el Tipo de Cambio Real (TC) y el Consumo de Gobierno (CG):

$$\hat{X} = \hat{a} + \hat{b}_1 PIB + \hat{b}_2 FBKF + \hat{b}_3 INPC + \hat{b}_4 tTC + \hat{b}_5 CG + \hat{e}$$

Para la obtención del Cuadro 12 se realizó la regresión escribiendo en la barra de comandos de EViews 5: **LS X C PIB FBKF INPC TC CG**

Cuadro 12. Resultados de la regresión.

Dependent Variable: X
Method: Least Squares
Modelo Original.
Sample: 1980Q1 2010Q1
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3382.124	2822.119	1.198434	0.2331
PIB	6.31E-06	5.61E-07	11.24808	0.0000
FBKF	3.68E-06	2.48E-06	1.482472	0.1409
INPC	-124.1930	113.8400	-1.090944	0.2775
TC	199.8320	430.3857	0.464309	0.6433
CG	2.95E-06	2.83E-06	1.044073	0.2986

R-squared	0.983829	Mean dependent var	33392.56
Adjusted R-squared	0.983143	S.D. dependent var	26035.35
S.E. of regression	3380.249	Akaike info criterion	19.13646
Sum squared resid	1.35E+09	Schwarz criterion	19.27293
Log likelihood	-1180.461	F-statistic	1435.767
Durbin-Watson stat	0.390360	Prob(F-statistic)	0.000000

Pruebas para detectar multicolinealidad.

1. Coeficiente de determinación R^2 alto y T´s no significativas.

Ahora bien, como se recordará, para determinar si los coeficientes estimados son estadísticamente significativos, es decir, que su valor sea distinto de cero, nos basamos en sus probabilidades asociadas. Para ello se establecen los siguientes criterios:

- Si la probabilidad asociada al coeficiente es menor o igual a 0.05 se dice que éste es estadísticamente significativo, por lo que, la variable independiente explica a la dependiente.

- Si la probabilidad asociada al coeficiente es mayor a 0.05 se dice que éste es estadísticamente no significativo, por lo que, la variable independiente no explica a la dependiente.

Por lo que respecta al estadístico F, planteamos los siguientes criterios de decisión:

- Si la probabilidad asociada al estadístico F es menor o igual a 0.05, se concluye que en su conjunto todas las variables independientes están explicando a la variable dependiente.

- Si la probabilidad asociada al estadístico F es mayor a 0.05, se concluye que en su conjunto todas las variables independientes no están explicando a la variable dependiente.

En nuestro caso, observamos (ver Cuadro 12) que el R^2 es alto, mientras que el estadístico F nos indica que en su conjunto todas las variables independientes explican a la variable dependiente. Sin embargo, las variables FBKF, INPC, TC y CG son estadísticamente no significativas ya que su probabilidad es mayor a 0.05; mientras que la probabilidad asociada a la variable PIB es menor a 0.05, indicando que ésta explica a la variable dependiente, es decir, es estadísticamente significativa. Estos resultados indican la posibilidad de alta multicolinealidad.

2. Matriz de correlaciones.

Para obtener la matriz de correlación se escribe en la barra de comandos de Eviews 5: **COR PIB FBKF INPC TC CG**

Cuadro 13. Matriz de correlaciones.

	PIB	FBKF	INPC	TC	CG
PIB	1.000000	-0.337159	0.972901	0.918392	0.683229
FBKF	-0.337159	1.000000	-0.385403	-0.449370	-0.424270
INPC	0.972901	-0.385403	1.000000	0.973883	0.684405
TC	0.918392	-0.449370	0.973883	1.000000	0.684609
CG	0.683229	-0.424270	0.684405	0.684609	1.000000

Para determinar si existe alta multicolinealidad por medio de la matriz de correlaciones se sugiere observar el coeficiente de correlación entre dos regresoras, si alguno de estos es mayor al coeficiente de determinación de la regresión original (R^2), entonces tenemos presencia de alta multicolinealidad.

Planteamos la siguiente prueba de hipótesis.

Ho: No multicolinealidad.
Ha: Multicolinealidad.

Regla de decisión:

Se acepta Ho si $R^2 > r_{x1}, r_{x2}, \ldots, r_{xn}$
Se acepta Ha si $R^2 < r_{x1}, r_{x2}, \ldots, r_{xn}$

Así, como ningún valor de la matriz de correlaciones es mayor al R^2 de la regresión original (Cuadro 13), se acepta la hipótesis nula de no multicolinealidad.

3. **Regresiones auxiliares (Prueba de Klein).**

Las regresiones auxiliares se introducirán de la siguiente forma en la barra de comando de Eviews 5.

Regresión auxiliar 1: **LS PIB C FBKF INPC TC CG**

Regresión auxiliar 2: **LS FBKF C INPC TC CG PIB**

Regresión auxiliar 3: **LS INPC C TC CG PIB FBKF**

Regresión auxiliar 4: **LS TC C CG PIB FBKF INPC**

Regresión auxiliar 5: **LS CG C PIB FBKF INPC TC**

En los Cuadros 14, 15, 16, 17 y 18 se muestra los resultados de las regresiones auxiliares.

Cuadro 14. Resultados de la Regresión auxiliar 1.

Dependent Variable: PIB
Method: Least Squares
Regresión Auxiliar 1
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3757.934	3226.954	1.164545	0.2465
FBKF	1.01E-05	4.77E-05	0.210911	0.8333
INPC	1153.259	61.66372	18.70347	0.0000
TC	-3310.687	467.1911	-7.086146	0.0000
CG	9.41E-06	4.29E-06	2.194375	0.0302

R-squared	0.964420	Mean dependent var	35609.22
Adjusted R-squared	0.963227	S.D. dependent var	26383.26
S.E. of regression	5059.301	Akaike info criterion	19.93533
Sum squared resid	3.05E+09	Schwarz criterion	20.04905
Log likelihood	1230.991	F-statistic	806.4705
Durbin-Watson stat	0.355470	Prob(F-statistic)	0.000000

Cuadro 15. Resultado de la Regresión auxiliar 2.

Dependent Variable: FBKF
Method: Least Squares
Regresión Auxiliar 2
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	25645589	5775870	4.440091	0.0000
INPC	322440.1	234077.7	1.377492	0.1709
TC	-3196879.	1029614.	-3.104931	0.0024
CG	-0.019679	0.008210	-2.396813	0.0181
PIB	37.14459	176.1151	0.210911	0.8333

R-squared	0.289575	Mean dependent var	4073817.
Adjusted R-squared	0.265695	S.D. dependent var	11344987
S.E. of regression	9721692.	Akaike info criterion	35.05710
Sum squared resid	1.12E+16	Schwarz criterion	35.17083
Log likelihood	-2168.540	F-statistic	12.12633
Durbin-Watson stat	0.243508	Prob(F-statistic)	0.000000

Cuadro 16. Resultados de la Regresión auxiliar 3.

Dependent Variable: INPC
Method: Least Squares
Regresión Auxiliar 3
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-6.233045	2.353337	-2.671970	0.0086
TC	3.960155	0.203002	19.50792	0.0000
CG	-3.25E-09	3.25E-09	-1.000517	0.3191
PIB	0.000646	3.44E-05	18.78347	0.0000
FBKF	4.87E-08	3.53E-08	1.377492	0.1709

R-squared	0.938174	Mean dependent var	36.06161
Adjusted R-squared	0.937777	S.D. dependent var	34.16457
S.E. of regression	3.777228	Akaike info criterion	5.535345
Sum squared resid	1697.826	Schwarz criterion	5.649066
Log likelihood	-338.1914	F-statistic	2465.903
Durbin-Watson stat	0.314359	Prob(F-statistic)	0.000000

Cuadro 17. Resultado de la Regresión auxiliar 4.

Dependent Variable: TC
Method: Least Squares
Regresión Auxiliar 4
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coeficient	Std. Error	t-Statistic	Prob.
C	1.145743	0.523569	2.188330	0.0306
CG	1.02E-09	7.14E-10	1.429936	0.1554
PIB	-8.96E-05	1.26E-05	-7.086146	0.0000
FBKF	-2.34E-08	7.55E-09	-3.104031	0.0024
INPC	0.192364	0.009861	19.50792	0.0000

R-squared	0.968296	Mean dependent var	6.005565
Adjusted R-squared	0.967231	S.D. dependent var	4.596804
S.E. of regression	0.832489	Akaike info criterion	2.510694
Sum squared resid	82.47153	Schwarz criterion	2.624415
Log likelihood	-150.6530	F-statistic	908.6282
Durbin-Watson stat	0.274739	Prob(F-statistic)	0.000000

Cuadro 18. Resultado de la Regresión auxiliar 5

Dependent Variable: CG
Method: Least Squares
Regresión Auxiliar 5
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.72E+08	28821469	23.31275	0.0000
PIB	4132.386	1883.172	2.194375	0.0302
FBKF	-2.340178	0.976371	-2.396813	0.0181
INPC	-2563458.	2562135	-1.000517	0.3191
TC	16551328	11574871	1.429936	0.1554

R-squared	0.518524	Mean dependent var	8.11E+08
Adjusted R-squared	0.502340	S.D. dependent var	1.50E+08
S.E. of regression	1.06E+08	Akaike info criterion	39.83555
Sum squared resid	1.34E+18	Schwarz criterion	39.94927
Log likelihood	-2464.804	F-statistic	32.03911
Durbin-Watson stat	3.288690	Prob(F-statistic)	0.000000

La regla de decisión será la siguiente:

- Si al menos un R^2 de una regresión auxiliar es mayor al R^2 de la regresión original, se dice que existe alta multicolinealidad.

- Si en ninguna de las regresiones auxiliares el R^2 es mayor al R^2 de la regresión original, se dice que no existe alta multicolinealidad.

Así, en base a los resultados de las regresiones auxiliares, se puede observar en la regresión auxiliar 3, el R^2 es mayor al R^2 de la regresión original, concluyéndose que existe alta multicolinealidad.

4. Prueba de Theil. Se aplicará la fórmula Theil, explicada en el ejercicio1: Consumo Gubernamental.

Para evaluar los resultados mediante esta prueba es necesario formular la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: Multicolinealidad.

La regla de decisión es:

- Si Theil ≤ a 0.5 se acepta la Ho.
- Si Theil > a 0.5 no se acepta la Ho.

La fórmula para determinar la prueba es la siguiente:

Theil= R^2-∑(R^2-R^2_i).

Donde:

R^2: coeficiente de determinación de la regresión original.
R^2_i: coeficiente de determinación de las regresiones auxiliares.

Se resta el R^2 del modelo original de cada una de las regresiones auxiliares.

R_i^2 :del modelo original=0.983829

R_j^2 :de la regresión auxiliar 1=0.964423

R_j^2 :de la regresión auxiliar 2=0.289575

R_j^2 :de la regresión auxiliar 3=0.988174

R_j^2 :de la regresión auxiliar 4=0.968296

R_j^2 :de la regresión auxiliar 5=0.518524

Así, sustituyendo los valores en la formula, obtenemos:

Theil = (0.983829 -∑ ((0.983829 - 0.964423) + (0.983829 - 0.289575) + (0.983829 - 0.988174) + (0.983829 - 0.968296) + (0.983829-0.518524)) = **0.2063**.

Tomando el resultado en valor absoluto, tenemos que como Theil es igual a 0.2063, es decir, menor a 0.5, entonces aceptamos la hipótesis nula de no multicolinealidad.

5. Factor de Inflación de varianza (FIV) o Regla de Klein.

Entonces en base a los resultados obtenidos de las regresiones auxiliares que se realizaron con anterioridad, calcularemos el FIV.

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: multicolinealidad.

La Regla de decisión es:

- Si el valor de FIV es ≤ a 10 se acepta Ho.
- Si el valor de FIV es > a 10 no se acepta Ho.

La fórmula es:

$$FIV = \frac{1}{(1-R^2)}$$

Determinación de multicolinealidad de manera conjunta:

$$FIV\ de\ la\ regresión\ original\ (Cuadro\ 14) = \frac{1}{(1-0.983829)}$$

$$FIV = \mathbf{61.839094.}$$

Con este resultado se observa que el modelo en conjunto presenta multicolinealidad, ya que el valor del FIV es mayor que 10.

Ahora estimaremos el FIV para cada una de las variables independientes para determinar cuál o cuáles son las variables independientes causantes de la alta multicolinealidad. Para ello se tomarán los R^2 correspondientes de las ecuaciones auxiliares previamente realizadas.

$$FIVreg.auxiliar\ 1 = \frac{1}{(1-0.964423)}$$

$$\boldsymbol{FIV = 28.10}$$

$$FIVreg.auxiliar\ 2 = \frac{1}{(1-0.289575)}$$

$$FIV = 1.40$$

$$FIVreg.auxiliar\ 3 = \frac{1}{(1-0.988174)}$$

$$\boldsymbol{FIV = 84.55}$$

$$FIVreg.auxiliar\ 4 = \frac{1}{(1-0.968296)}$$

$$\boldsymbol{FIV = 31.54}$$

$$FIVreg.auxiliar\ 5 = \frac{1}{(1-0.518524)}$$

$$FIV = 2.07$$

Los FIV de las regresiones auxiliares 1, 3 y 4 son mayores a 10, por lo que se acepta la hipótesis de multicolinealidad; mientras que el FIV de las regresiones auxiliares 2 y 5, son menores a 10 por lo que se acepta la hipótesis de no multicolinealidad.

Prueba 6. Prueba de Farrar y Glouber.

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: multicolinealidad.

La regla de decisión es:

- Si el valor de X^2 es mayor o igual al valor de G se acepta la Ho.
- Si el valor de X^2 es menor al valor de G no se acepta la Ho.

La G denota el valor del estadístico construido por Farrar y Glouber.

La fórmula de esta prueba es la siguiente**:**

$$G = -\left(N - \frac{1}{6}(2K+5)\right) In\ |Rxx|$$

$$G \sim x^2$$

Grados de libertad = $n = k * \frac{k-1}{2}$

Donde:

$|Rxx|$ = Matriz de correlación de las variables independientes del valor absoluto.

N= número de observaciones.

K=Variables independientes incluyendo la constante.

Grados de libertad= $n = 6 * \frac{6-1}{2}$= 15

Valor de X^2=24.99

Como son conocidos todos los valores, puede calcularse sustituyendo los valores en la formula:

$$G = -\left(124 - \frac{1}{6}((2*6)+5)\right) In \left| 0.00068463 \right| = 882.8961$$

Como X^2= 24.99 < G =882.8961, entonces se concluye que existe multicolinealidad.

Finalmente para este ejemplo se puede decir que el grado de correlación entre las variables independientes es alto ya que se demostró mediante diversas pruebas.

Se corregirá el problema de multicolinealidad con la transformación a logaritmos de las variables, PIB y FBKF y con la introducción de la variable M (importaciones de bienes y servicios). Y eliminando TC y CG. En Eviews 5 se escribe:

Ls log(X) c log(PIB) log(FBKF) INPC M

Como las variables no se habían transformando a logaritmos se procede a escribir en la barra de comandos **log(X) log(PIB) log(FBKF),** para especificar la transformación directamente.

Cuadro 19. Resultados de la regresión.

Dependent Variable: LOG(X)
Method: Least Squares
Modelo Corregido
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.340520	0.433251	14.63443	0.0000
LOG(PIB)	0.403011	0.040405	10.04132	0.0000
LOG(FBKF)	-0.070020	0.007366	-9.506351	0.0000
INPC	0.008977	0.001179	7.612330	0.0000
M	4.95E-06	1.46E-06	3.388330	0.0010

R-squared	0.990220	Mean dependent var	10.73332
Adjusted R-squared	0.989901	S.D. dependent var	0.833763
S.E. of regression	0.086846	Akaike info criterion	-2.006872
Sum squared resid	0.897526	Schwarz criterion	-1.893151
Log likelihood	129.3121	F-statistic	3012.072
Durbin-Watson stat	0.395202	Prob(F-statistic)	0.000000

Así, se observa que los coeficientes son estadísticamente significativos ya que sus probabilidades asociadas son menores a 0.05. Asimismo, la probabilidad asociada al estadístico F es menor a 0.05, indicando que las variables independientes explican en su conjunto a la variable dependiente. Por lo que respecta al R^2, su valor es alto. Por tanto, concluimos que el modelo no presenta multicolinealidad.

En base al criterio que se estableció en el ejemplo 1, para fines prácticos, nos quedaremos con la prueba donde las variables independientes son estadísticamente significativas al lado de una prueba F significativa y un R^2 alto.

Ejemplo 3: Importaciones

Se realizara la regresion de las importaciones en función del Consumo Privado (CP), la Formación Bruta de Capital (FBKF) y el Consumo de Gobierno (CG).

$$\hat{M} = \hat{a} + \hat{b}_1 CP + \hat{b}_2 FBKF + \hat{b}_3 CG + \hat{e}$$

Para la obtención del Cuadro 20 se realizó la regresión escribiendo en la barra de comandos:**LS M C CP FBKF CG.**

Cuadro 20. Resultados de la regresión.

Dependent Variable: M
Method: Least Squares
Regresión Original.
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-61173.23	2205.992	-27.73048	0.0000
CP	1.98E-05	1.35E-06	14.69184	0.0000
FBKF	1.14E-05	3.30E-06	3.454740	0.0008
CG	-8.87E-07	3.47E-06	-0.255585	0.7987

R-squared	0.976706	Mean dependent var	35609.22
Adjusted R-squared	0.976123	S.D. dependent var	26383.26
S.E. of regression	4076.775	Akaike info criterion	19.49573
Sum squared resid	1.99E+09	Schwarz criterion	19.58670
Log likelihood	-1204.735	F-statistic	1677.146
Durbin-Watson stat	0.730435	Prob(F-statistic)	0.000000

1. Coeficiente de determinación R^2 alto y T´s no significativas.

Para determinar si los coeficientes estimados son estadísticamente significativos, es decir, que su valor sea distinto de cero, nos basamos en sus probabilidades asociadas. Para ello se establecen los siguientes criterios:

- Si la probabilidad asociada al coeficiente es menor o igual a 0.05 se dice que éste es estadísticamente significativo, por lo que, la variable independiente explica a la dependiente.

- Si la probabilidad asociada al coeficiente es mayor a 0.05 se dice que éste es estadísticamente no significativo, por lo que, la variable independiente no explica a la dependiente.

Por lo que respecta al estadístico F, planteamos los siguientes criterios de decisión:

- Si la probabilidad asociada al estadístico F es menor o igual a 0.05, se concluye que en su conjunto todas las variables independientes están explicando a la variable dependiente.

- Si la probabilidad asociada al estadístico F es mayor a 0.05, se concluye que en su conjunto todas las variables independientes no están explicando a la variable dependiente.

Como se puede observar en el Cuadro 20 en valor de la R^2 es alto, mientras que el estadístico F nos indica que en su conjunto todas las variables independientes explican a la variable dependiente. Sin embargo, las variables CP y FBKF son estadísticamente significativas ya que su probabilidad es menor a 0.05; mientras que la probabilidad asociada a la variable CG es mayor a 0.05, indicando que ésta no explica a la variable dependiente, es decir, es no estadísticamente significativa. Estos resultados indican la posibilidad de alta multicolinealidad.

2. Matriz de correlaciones.

Para determinar si existe alta multicolinealidad por medio de la matriz de correlaciones se sugiere observar el coeficiente de correlación entre dos regresoras, si alguno de estos es mayor al coeficiente de determinación de la regresión original (R^2), entonces tenemos presencia de alta multicolinealidad.

Planteamos la siguiente prueba de hipótesis.

Ho: No multicolinealidad.
Ha: Multicolinealidad.

Regla de decisión:

Se acepta Ho si $R^2 > r_{x1}, r_{x2}, \ldots, r_{xn}$
Se acepta Ha si $R^2 < r_{x1}, r_{x2}, \ldots, r_{xn}$

Para obtener la matriz de correlación en Eviews en la barra de comandos se escribe: **COR CP FBKF CG**, lo cual muestra la matriz de correlaciones, como en el Cuadro 21.

Cuadro 21. Matriz de correlaciones.

	CP	FBKF	CG
CP	1.000000	0.962360	0.700360
FBKF	0.962360	1.000000	0.643417
CG	0.700360	0.643417	1.000000

Así, como los valores de la matriz de correlación son mayores que el R^2 del Cuadro 22, no se rechaza la hipótesis nula de no multicolinealidad.

3. Regresiones auxiliares (Prueba de Klein).

Se realizarán tres regresiones auxiliares, en donde cada variable independiente tomará el papel, ahora, de variable dependiente, con la finalidad de comparar los R^2 de cada una de ellas respecto al R^2 de la regresión original.

La regla de decisión será la siguiente:

- Si al menos un R^2 de una regresión auxiliar es mayor al R^2 de la regresión original, se dice que existe alta multicolinealidad.
- Si en ninguna de las regresiones auxiliares el R^2 es mayor al R^2 de la regresión original, se dice que no existe alta multicolinealidad.

Dentro de Eviews, en la barra de comandos, se escribe: **LS CP C FBKF CG** para obtener Cuadro 22.

Cuadro 22. Resultados de la Regresión auxiliar 1

Dependent Variable: CP
Method: Least Squares
Regresión auxiliar 1.
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.29E+08	1.37E+08	4.592354	0.0000
FBKF	2.239902	0.077929	29.38452	0.0000
CG	1.002168	0.215040	4.660374	0.0000

R-squared	0.937377	Mean dependent var	4.22E+09
Adjusted R-squared	0.936342	S.D. dependent var	1.09E+09
S.E. of regression	2.74E+08	Akaike info criterion	41.72171
Sum squared resid	9.11E+18	Schwarz criterion	41.78994
Log likelihood	-2583.746	F-statistic	905.6009
Durbin-Watson stat	0.650213	Prob(F-statistic)	0.000000

Para obtener la segunda regresión auxiliar se escribe en la barra de comando en Eviews: **LS FBKF C CG CP** (obteniendo Cuadro 23).

Cuadro 23. Resultados de la Regresión auxiliar 2

Dependent Variable: FBKF
Method: Least Squares
Regresión Auxiliar 2.
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.69E+08	55571689	-4.842534	0.0000
CG	-0.165630	0.094321	-1.756020	0.0816
CP	0.383025	0.013035	29.38452	0.0000

R-squared	0.927972	Mean dependent var	1.21E+09
Adjusted R-squared	0.926782	S.D. dependent var	4.15E+08
S.E. of regression	1.12E+08	Akaike info criterion	39.93354
Sum squared resid	1.52E+18	Schwarz criterion	40.00178
Log likelihood	-2472.880	F-statistic	779.4527
Durbin-Watson stat	0.322620	Prob(F-statistic)	0.000000

Con la tercera regresión auxiliar se realizará lo mismo que las anteriores, sólo el comando cambia: **LS CG C FBK CP** (obteniendo Cuadro 24)

Cuadro 24. Resultados de la Regresión auxiliar 3

Dependent Variable: CG
Method: Least Squares
Regresión auxiliar 3.
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.53E+08	48069375	7.340759	0.0000
FBKF	-0.150040	0.085443	-1.756020	0.0816
CP	0.151851	0.032584	4.660374	0.0000

R-squared	0.503165	Mean dependent var	8.11E+08
Adjusted R-squared	0.494953	S.D. dependent var	1.50E+08
S.E. of regression	1.07E+08	Akaike info criterion	39.83469
Sum squared resid	1.38E+18	Schwarz criterion	39.90292
Log likelihood	-2466.751	F-statistic	61.27081
Durbin-Watson stat	3.002056	Prob(F-statistic)	0.000000

Así, en base a los resultados de las regresiones auxiliares, se puede observar que tanto en la regresión auxiliar 1, 2 y 3 el R^2 es menor al R^2 de la regresión original, concluyéndose que no existe multicolinealidad.

Prueba 4. Prueba de Theil.

Para evaluar los resultados mediante esta prueba es necesario formular la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: Multicolinealidad.

La regla de decisión es:

- Si Theil ≤ a 0.5 se acepta la Ho.
- Si Theil > a 0.5 no se acepta la Ho.

La fórmula para determinar la prueba es la siguiente:

Theil= R^2-∑(R^2-R^2_i).

Donde:

R^2: coeficiente de determinación de la regresión original.
R^2_i: coeficiente de determinación de las regresiones auxiliares.

Se resta el R^2 del modelo original de cada una de las regresiones auxiliares.

La fórmula de la prueba de Theil es:

Theil= R^2-∑(R^2-R^2_i).

R^2 original = 0.976706

R^2 regresión auxiliar 1= 0.937377

R^2 regresión auxiliar 2= 0.927972

R^2 regresión auxiliar 3= 0.503165

Theil = (0.976706 - ∑[(0.976706 - 0.937377) + (0.976706 - 0.927972) + (0.976706 - 0.503165)] = **0.415102.**

Tomando el resultado en valor absoluto, tenemos que como Theil es igual a 0.415102, es decir, menor a 0.5, entonces aceptamos la hipótesis nula de no multicolinealidad.

5. Factor de Inflación de Varianza (FIV)

Basado en las regresiones auxiliares, el FIV, muestra como la varianza de un estimador se infla por la presencia de multicolinealidad. Si lo que se quiere es determinar de manera conjunta si el modelo presenta multicolinealidad, entonces, se calcula el FIV para la regresión original. Asimismo, si buscamos determinar que variables es causante de multicolinealidad, entonces se utilizan las regresiones auxiliares, para el cálculo del FIV.

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: multicolinealidad.

La Regla de decisión es:

- Si el valor de FIV es ≤ a 10 se acepta Ho.
- Si el valor de FIV es > a 10 no se acepta Ho.

La fórmula es:

$$FIV = \frac{1}{(1 - R^2)}$$

Determinación de multicolinealidad de manera conjunta:

$$FIV = \frac{1}{(1 - 0.976706)} = 42.9295$$

Así, como el valor del FIV es de 42.9295, entonces no aceptamos Ho. Por tanto, se concluye que, de acuerdo a esta prueba, existe alta multicolinealidad en el modelo.

Ahora estimaremos el FIV para cada una de las variables independientes para determinar cuál o cuáles son las variables independientes causantes de la alta multicolinealidad. Para ello se tomarán los R^2 correspondientes de las ecuaciones auxiliares previamente realizadas.

R^2 regresión auxiliar 1= -0.937377

R^2 regresión auxiliar 2= 0.927972

R^2 regresión auxiliar 3= 0.503165

A continuación se muestran los valores del FIV para cada regresión auxiliar:

$$FIVreg.auxiliar1 = \frac{1}{(1 - 0.937377)} = 0.062623$$

$$FIV = 0.062623$$

$$FIVreg.auxiliar2 = \frac{1}{(1 - 0.927372)} = 0.072028$$

$$FIV = 0.072028$$

$$FIVreg.auxiliar3 = \frac{1}{(1-0.503165)} = 0.496835$$

$$FIV = 0.496835$$

Los factores de inflación de varianza de la regresiones auxiliares 1, 2 Y 3 son menores a 10, por lo que se acepta la hipótesis de no multicolinealidad.

7. **Prueba Farrar- Glouber.**

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: multicolinealidad.

La regla de decisión es:

- Si el valor de X^2 es mayor o igual al valor de G se acepta la Ho.
- Si el valor de X^2 es menor al valor de G no se acepta la Ho.

La G denota el valor del estadístico construido por Farrar y Glouber.

La fórmula de esta prueba es la siguiente**:**

$$G = -\left(N - \frac{1}{6}(2K+5)\right) In\ |Rxx|$$

$$G \sim x^2$$

Grados de libertad = $n = k * \frac{k-1}{2}$

Donde:

$|Rxx|$ = Matriz de correlación de las variables independientes del valor absoluto.

N= número de observaciones.

K=Variables independientes incluyendo la constante.

Sustituyendo:

Grados de libertad= $4 * \frac{4-1}{2}$= 6

X^2 con 6 grados de libertad (gl) = 12.59

$$ln|0.036699795| = -3.30503447$$

$$G = -\left(124 - \frac{1}{6}(2*(4)+5)\right)* -3.30503447$$

$$G = -131.160908$$

Así, como el valor de $X^2 = 12.59 < G = -131.160908$, no se acepta la hipótesis nula de no multicolinealidad.

Por lo tanto el modelo presenta multicolinealidad.

Como se ha observado, en los contrastes de Theil, Farrar- Glouber y en la prueba de R^2 alto y T‘s, se puede concluir que hay presencia de multicolinealidad.

Corrección de multicolinealidad.

Se transformaran las variables FBKF y CG a logaritmos.

El nuevo modelo planteado es el siguiente:

$$L\hat{M} = \hat{a} + \hat{b}CP + \hat{c}LFBJF + \hat{d} * LCG + \hat{e}$$

En la barra de comandos de Eviews 5 se escribe **LS LM C CP LFBKF LCG** y de esa forma se obtiene el Cuadro 25.

Cuadro 25. Resultados de la regresión.

Dependent Variable: LM
Method: Least Squares
Medida de corrección con logaritmos 1.
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-15.64577	5.419992	-2.886678	0.0046
CP	4.65E-10	6.66E-11	6.978885	0.0000
LFBKF	0.773022	0.195229	3.959574	0.0001
LCG	0.376710	0.168944	2.229797	0.0276

R-squared	0.913072	Mean dependent var	10.16112
Adjusted R-squared	0.910899	S.D. dependent var	0.845756
S.E. of regression	0.252457	Akaike info criterion	0.116572
Sum squared resid	7.648124	Schwarz criterion	0.207549
Log likelihood	-3.227469	F-statistic	420.1507
Durbin-Watson stat	0.435203	Prob(F-statistic)	0.000000

1. **Coeficiente de determinación R^2 alto y T´s no significativas**.

Así, se observa que los coeficientes son estadísticamente significativos ya que sus probabilidades asociadas son menores a 0.05. Asimismo, la probabilidad asociada al estadístico F es menor a 0.05, indicando que las variables independientes explican en su conjunto a la variable dependiente. Por lo que respecta al R^2, su valor es alto. Por tanto, concluimos que el modelo no presenta multicolinealidad.

2. Matriz de correlaciones.

Para obtener la matriz de correlación, se escribe en la barra de comandos **COR CP LFBKF LCG.**

Cuadro 26. Matriz de correlaciones.

	CP	LFBKF	LCG
CP	1.000000	0.938712	0.690608
LFBKF	0.938712	1.000000	0.616563
LCG	0.690608	0.616563	1.000000

Para determinar si existe multicolinealidad por medio de la matriz de correlaciones planteamos la siguiente prueba de hipótesis.

Ho: No multicolinealidad
Ha: Multicolinealidad

Regla de decisión:

Se acepta Ho si $R^2 > r_{x1}, r_{x2},\ldots, r_{xn}$
Se acepta Ha si $R^2 < r_{x1}, r_{x2},\ldots, r_{xn}$

Por tanto, como el coeficiente de correlación entre las variables independientes LFBKF vs CP es mayor que el R^2, no se acepta la hipótesis nula de no multicolinealidad.

3. Regresiones Auxiliares (Prueba de Klein).

Las regresiones auxiliares para este caso son:

Regresión auxiliar 1**: IS CP C LFBK LCG**

Regresión auxiliar 2: **Is CP c LFBKF LCP**

Regresión auxiliar 3: **IS LCG C LFBK CP**.

Los resultados se muestran a continuación:

Cuadro 27. Resultado de la Regresiones auxiliar 1.

Dependent Variable: CP
Method: Least Squares

Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-103324.8	5149.423	-20.06532	0.0000
LFBK	50402.24	4857.265	10.37667	0.0000
LCG	-33412.20	4699.112	-7.110322	0.0000

R-squared	0.821099	Mean dependent var	12527.98
Adjusted R-squared	0.818142	S.D. dependent var	32591.44
S.E. of regression	13898.56	Akaike info criterion	21.94085
Sum squared resid	2.34E+10	Schwarz criterion	22.00909
Log likelihood	1357.333	F-statistic	277.6759
Durbin-Watson stat	0.422844	Prob(F-statistic)	0.000000

Cuadro 28. Resultados de la Regresión auxiliar 2.

Dependent Variable: LFBK
Method: Least Squares

Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.288052	0.086933	14.81656	0.0000
CP	9.34E-06	9.00E-07	10.37667	0.0000
LCG	0.818341	0.016384	49.94863	0.0000

R-squared	0.988267	Mean dependent var	6.109253
Adjusted R-squared	0.988073	S.D. dependent var	1.732631
S.E. of regression	0.189221	Akaike info criterion	-0.467910
Sum squared resid	4.332336	Schwarz criterion	-0.399678
Log likelihood	32.01045	F-statistic	5095.946
Durbin-Watson stat	0.972897	Prob(F-statistic)	0.000000

Cuadro 29. Resultados de la Regresión auxiliar 3.

Dependent Variable: LCG
Method: Least Squares

Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.261182	0.130940	-9.631776	0.0000
LFDK	1.165461	0.023333	49.94863	0.0000
CP	-8.82E-06	1.24E-06	-7.110322	0.0000

R-squared	0.984361	Mean dependent var	5.748417
Adjusted R-squared	0.984102	S.D. dependent var	1.790944
S.E. of regression	0.225814	Akaike info criterion	-0.114317
Sum squared resid	6.170008	Schwarz criterion	-0.046085
Log likelihood	10.08767	F-statistic	3807.967
Durbin-Watson stat	0.990867	Prob(F-statistic)	0.000000

La regla de decisión es la siguiente:

- Si al menos un R^2 de una regresión auxiliar es mayor al R^2 de la regresión original, se dice que existe alta multicolinealidad.
- Si en ninguna de las regresiones auxiliares el R^2 es mayor al R^2 de la regresión original, se dice que no existe alta multicolinealidad.

Así, dado que los R^2 de dos regresiones auxiliares son mayores que el R^2 del nuevo modelo, se concluye, mediante esta prueba, que el modelo presenta alta multicolinealidad.

4. Prueba de Theil.

Para evaluar los resultados mediante esta prueba es necesario formular la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: Multicolinealidad.

La regla de decisión es:

- Si Theil ≤ a 0.5 se acepta la Ho.
- Si Theil > a 0.5 no se acepta la Ho.

La fórmula para determinar la prueba es la siguiente:

Theil= $R^2-\sum(R^2-R^2_i)$.

Donde:

R^2: coeficiente de determinación de la regresión original.
$R^2_{i:}$ coeficiente de determinación de las regresiones auxiliares.

Se resta el R^2 del modelo original de cada una de las regresiones auxiliares. Así:

R^2 original = 0.913072

R^2 regresión auxiliar 1= 0.821099

R^2 regresión auxiliar 2= 0.988267

R^2 regresión auxiliar 3= 0.984361

Theil: 0.883101-∑[(0.883101-0.821099)+(0.883101-0.988267)+(0.883101-0.984361)]=**-0.261323.**

Tomando el resultado en valor absoluto, tenemos que Theil es igual a 0.670566, como es menor a 0.5 aceptamos la hipótesis de que no existe alta multicolinealidad.

5. Prueba de Factor de Inflación de Varianza (FIV)

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: multicolinealidad.

La Regla de decisión es:

- Si el valor de FIV es ≤ a 10 se acepta Ho.
- Si el valor de FIV es > a 10 no se acepta Ho.

La fórmula es:

$$FIV = \frac{1}{(1-R^2)}$$

Los resultados son los siguientes:

R^2 original = 0.913072

R^2 regresión auxiliar 1= 0.821099

R^2 regresión auxiliar 2= 0.988267

R^2 regresión auxiliar 3= 0.984361

Sustituyendo el valor de R^2 del modelo de corrección con logaritmos 1.

$$FIVreg.original = \frac{1}{(1-0.913072)} = 11.50377$$

Así, como el valor del FIV es mayor que 10, se no acepta la hipótesis nula de no multicolinealidad.

Ahora estimaremos el FIV para cada una de las variables independientes para determinar cuál o cuáles son las variables independientes causantes de la alta multicolinealidad. Para ello se tomarán los R^2 correspondientes de las ecuaciones auxiliares previamente realizadas.

A continuación se muestran los valores del FIV para cada regresión auxiliar:

R^2 regresión auxiliar 1= 0.821099

R^2 regresión auxiliar 2= 0.988267

R^2 regresión auxiliar 3= 0.984361

$$FIVreg.auxiliar1 = \frac{1}{(1-0.821099)} = 5.589683$$

$$FIVreg.auxiliar2 = \frac{1}{(1-0.988267)} = 85.229694$$

$$FIVreg.auxiliar2 = \frac{1}{(1-0.984361)} = 63.94$$

De acuerdo con el Factor de Inflación de Varianza, los valores que presentan las regresiones auxiliares 2 y 3 son mayores a 10, entonces no se acepta la hipótesis nula de no multicolinealidad.

6. Prueba Farrar- Glouber.

Se establece la siguiente prueba de hipótesis:

Ho: No multicolinealidad.
Ha: Multicolinealidad.

La regla de decisión es:

- Si el valor de X^2 es mayor o igual al valor de G se acepta la Ho.
- Si el valor de X^2 es menor al valor de G no se acepta la Ho.

La G denota el valor del estadístico construido por Farrar y Glouber.
La fórmula de esta prueba es la siguiente**:**

$$G - -\left(N - \frac{1}{6}(2K + 5)\right) In\ |Rxx|$$

$$G \sim x^2$$

Grados de libertad = $n = k * \frac{k-1}{2}$

Donde:

$|Rxx|$ = Matriz de correlación de las variables independientes del valor absoluto.

N= número de observaciones.

K=Variables independientes incluyendo la constante.

$$G = -\left(N - \frac{1}{6}(2K + 5)\right) In\ |Rxx|$$

$$G \sim x^2$$

Sustituyendo:

$$G = \left(124 - \frac{1}{6}(2 * (4) + 5)\right) * -6.98690367$$

$$G = 11247.7504$$

Para los Grados de libertad:

Grados de libertad= $4 * \frac{4-1}{2}$= 6

X^2 con 6 grados de libertad (gl) = 12.59

En contraste de la prueba Farrar-Glouber con la X^2, se obtiene que:

Entonces X^2= 12.59 < G= 11247.7504, por lo que no se acepta la hipótesis nula de no multicolinealidad.

Como se desprende de los resultados obtenidos, algunas pruebas nos indican que la multicolinealidad disminuyó, sin embargo, algunas otras nos indican su alta persistencia. Por lo que, para fines prácticos, el estudioso o investigador que se encuentre en alguna situación similar, es decir, de contradicción entre los resultados de las pruebas, para el caso de multicolinealidad, nos quedaremos con la prueba donde las variables independientes son estadísticamente significativas al lado de una prueba F significativa y un R^2 alto.

VI.3.5.-REACTIVOS PARA REAFIRMAR LOS CONOCIMIENTOS

A.- AUTOEVALUACION: Repaso para reafirmar el conocimiento

Tema: Efecto sobre los estimadores de la violación de los supuestos del modelo de regresión lineal: Mínimos cuadrados ordinarios: multicolinealidad.

Conteste con una "X" en **SI** cuando la frase afirmativa sea verdadera y también con una "X" en NO cuando la frase afirmativa sea falsa:

1.- Existe cuando dos o más variables explicativas están altamente correlacionadas: SI___; NO____.

2.-Las consecuencias son que los estimadores pueden no ser estadísticamente significativos, sin embargo para propósitos de proyección eso no importa: SI____; NO______:

3.- Se identifica cuando se observa que alguno o ninguno de los coeficientes es estadísticamente significativo, R^2 es alta y F muestra que en conjunto si son significativos estadísticamente estos estimadores: SI____;NO_____.

4.- La Multicolinealidad también se identifica con la Prueba Theil y el Factor de Inflación de Varianza,(FIV) ? : Si___; NO___

5.- En opinión de algunos autores no es importante que exista la Multicolinealidad entre las variables explicativas cuando se va a proyectar: SI___; NO____

6.- Derivado de lo anterior podemos decir que l.a detección de multicolinealidad es un problema de la muestra no de la población; o sea que es una cuestión de grado y no de clase: SI____; NO______

7.- Para resolver el problema de multicolinealidad pueden: a).- ampliarse el tamaño de la muestra; b).- utilizar información a priori; c) Se transforma la relación funcional o d).- Se omite una de las variables altamente colineales: SI ____; NO_____.

8.- La multicolinealidad entre las variables explicativas de la ecuación de regresión también se elimina con la D de diferencias entre las variables explicativas: SI; _____; NO___ .

9.- Con el Método de deflactación se elimina la multicolinealidad y se corrige el problema de los signos de los coeficientes de las variables explicativas, cuando dichos signos no son congruentes con la teoría económica planteada: SI___; NO___.

10.- En opinión de Carrascal et al (2001:177) para corregir la multicolinealidad también existen el método de estimación conocido con el nombre de "Crestas", al igual que el método de componentes principales al conjunto de variables exógenas: SI___NO___.

VII.- EVALUACIÓN GENERAL DE LOS CONOCIMIENTOS BÁSICOS DE LA ECONOMETRÍA.

VII.1.- Primera parte: Conceptos básicos.

Indicaciones: Abunde y desarrolle teóricamente sobre cada uno de los siguientes temas:

1. Orígenes, primeros casos, evolución, desarrollo y perspectivas de la econometría.
2. Definición, objetivo, alcance, aplicaciones y limitaciones de la econometría.
3. Principales modelos econométricos desarrollados y utilizados en México.
4. Tipos de variables utilizadas en la econometría.
5. Tipos de datos utilizados en la econometría.
6. ¿Qué es un modelo? ¿Qué es una ecuación? ¿Qué es un parámetro?
7. Tipos y clasificación de modelos.
8. ¿Cuándo hacer modelos uniecuacionales o multiecuacionales para la estimación de los parámetros? ¿Cuáles se utilizan más en México y/o en la práctica y porqué?
9. Describa la metodología o pasos necesarios para plantear y solucionar un modelo econométrico.
10. Describa los diferentes métodos de estimación de los parámetros: MCO, Momentos, Máxima Verosimilitud y Residuos.
11. Describa los supuestos del método de Mínimos Cuadrados Ordinarios (MCO).
12. ¿Para qué sirve y cómo se obtiene el sistema de ecuaciones normales?
13. Definir ¿Qué se entiende por regresión, correlación y la relación y/o diferencia entre ellas?
14. ¿Qué es y para qué se utiliza el diagrama de dispersión?
15. Establezca las diferencias entre los Modelos Lineales Simples (MLS) y los Modelos Lineales Múltiples (MLM).

16. ¿Cuándo aplica o en qué casos la regresión y correlación no lineal en las variables?
17. ¿Qué son y qué papel desempeñan las U_i en las ecuaciones de regresión?
18. ¿Para qué sirve y cómo se utiliza la prueba de hipótesis? Establezca los pasos para plantear la Ho y la Ha.
19. ¿Qué es el error tipo I y II? ¿Qué es "α" y "ξ"?
20. ¿Para qué sirve el error estándar (s) de la regresión y de cada uno de los coeficientes de la regresión?
21. ¿Qué son y para qué sirven las bandas o intervalos de confianza?
22. ¿Establezca la diferencia e indique en que caso se utilizan "t", "z", "χ^2" y "F"?
23. Describa cuáles son e ilustre con literales y gráficas las propiedades de los estimadores.
24. Establezca la diferencia y similitud entre el coeficiente de correlación parcial y el múltiple.
25. Defina y establezca la relación y diferencia numérica entre "r", "R^2" y "R^2-ajustado".
26. ¿Qué son y para qué sirven las variables ficticias? ¿Qué es y para qué sirve el coeficiente de correlación de Spearman?
27. ¿Cómo se utiliza el modelo econométrico para hacer predicción y cómo se evalúa la calidad predictiva del modelo?
28. Explique ¿Cuándo y porqué se transforman las variables no lineales en lineales? ¿Cuáles son los métodos más comunes para transformar las variables no lineales en lineales? ¿Describa las características de los modelos: Lin-Lin, Log-Log, Lin-Log y Log-Lin, así como la forma funcional reciproca?¿Cómo se interpretan los parámetros obtenidos a través de cada uno de los modelos de la pregunta anterior? Incluya además un ejemplo del modelo reciproco para ver sus aplicaciones.
29. ¿Qué es heteroscedasticidad, autocorrelación y multicolinealidad? ¿Por qué surgen cada una de ellas?¿Qué consecuencias acarrean u ocasionan cada una de ellas en la ecuación de regresión? ¿Cuáles son los métodos para detectar y corregir cada una de ellas?
30. Describa los métodos para resolver un sistema de ecuaciones simultáneas tanto con información incompleta como completa

VII.2.-Segunda parte: Aplicación.

Indicaciones: Con base en la teoría de la primera parte, desarrolle el modelo asignado y realice lo que a continuación se le solicita.

Modelo "A":

PIB = f(Exportaciones, Importaciones, IED, Tipo de Cambio, INPC)

Modelo "B":

Exportaciones = f(PIB, Importaciones, IED, Tipo de Cambio, INPC)

1. Establezca el marco teórico: Planteamiento de la teoría económica que sustenta la relación entre las variables independientes con respecto a la dependiente, establezca la relación esperada entre cada una de las variables independientes con la dependiente (profundizar y revisar bibliografía al respecto). Asimismo, describa la naturaleza de los datos (su tipo) y realice la descripción del periodo histórico (1980:1 – 2009:2).

2. Planteamiento del modelo econométrico.

3. Resultados del modelo interpretando todos los coeficientes y demás medidas estadísticas: Escribir la ecuación de regresión; interpretar los coeficientes; interpretar los signos obtenidos y contrastarlos con los signos esperados, en caso de no coincidir los signos obtenidos con los signos esperados indicar a qué se debe o atribuye; interpretar las pruebas t's y F (cada una de ellas con su respectiva prueba de hipótesis, regla de decisión y conclusión); interpretar el coeficiente de determinación (R^2); la prueba DW; Criterio de Akaike y Shwarz; Asimetría, Kurtosis, Jarque-Bera; explique la diferencia entre el error estándar de la regresión y el de cada uno de los estimadores, construya e interprete estadísticamente las bandas de confianza con el error estándar de la regresión y con cada uno de los errores estándar de los estimadores.

   3.1. Pruebe, a través de los diferentes métodos informales y formales, que el modelo original (punto 3) presenta o no heteroscedasticidad analizando e interpretando cada uno de sus resultados. En caso de presentar esta anomalía, solucínela y describa el método que utilizó para resolverla, demostrando, a través de los diferentes métodos con su respectivo análisis e interpretación, que efectivamente la corrigió. (*Nota: Lo único que tiene que corregir en este punto es heteroscedasticidad, es decir, no importa que el modelo presente autocorrelación y/o multicolinealidad*).

   3.2. Pruebe, a través de los diferentes métodos informales y formales, que el modelo original (punto 3) presenta o no autocorrelación analizando e interpretando cada uno de sus resultados. En caso de presentar esta anomalía, solucínela y describa el método que utilizó para resolverla, demostrando, a través de los diferentes métodos con su respectivo análisis e interpretación, que efectivamente la corrigió. (*Nota: Lo único que tiene que corregir en este punto es autocorrelación, es decir, no importa que el modelo presente heteroscedasticidad y/o multicolinealidad*).

   3.3. Pruebe, a través de los diferentes métodos, que el modelo original (punto 3) presenta o no multicolinealidad analizando e interpretando cada uno de sus resultados. En caso de presentar esta anomalía, solucínela y describa el método que utilizó para resolverla, demostrando, a través de los diferentes métodos con su respectivo análisis e interpretación, que efectivamente la corrigió. (*Nota: Lo único que tiene que corregir en este punto es multicolinealidad, es decir, no importa que el modelo presente autocorrelación y/o heteroscedasticidad*).

4. Sobre el modelo del punto 3 (modelo original), obtenga *un* modelo exento de autocorrelación, heteroscedasticidad y multicolinealidad. Describa que procedimiento realizó y que métodos utilizó para obtenerlo y demuestre que efectivamente dicho modelo no presenta ninguna de estas anomalías a través de los distintos métodos conocidos, cada uno de ellos con su respectiva prueba de hipótesis, regla de decisión y conclusión. Además, no olvide realizar su análisis e interpretación correspondiente.

5. Conclusiones y recomendaciones.

6. Bibliografía General.

Observaciones:

1. El estudio que usted realice debe contener: Carátula, índice e introducción y debe estar engargolado.
2. Únicamente puede omitir, en caso de ser necesario, una variable independiente del modelo.
3. Se le recuerda que debe realizar el análisis e interpretación de cada uno de los resultados obtenidos, así como no omitir las respectivas pruebas de hipótesis, reglas de decisión y conclusión de cada una de las pruebas aplicadas.
4. Si incluye un anexo o los datos, éstos deben ir al final.

Base de datos:

Donde: PIB = Producto Interno Bruto.

XByS = Exportaciones de Bienes y Servicios.

MByS = Importaciones de Bienes y Servicios.

IED = Inversión Extranjera Directa.

TC = Tipo de Cambio.

INPC = Índice Nacional de Precios al Consumidor.

Datos de la Economía Mexicana, 1980:1-2009:2.						
Periodo	**PIB**	**XByS**	**MByS**	**IED**	**TC**	**INPC**
	Millones de pesos a precios corrientes	Millones de dólares	Millones de dólares	Millones de dólares	Pesos por dólar de EU	2002 = 100
1980/01	4332.88	5810.92	7146.48	400.81	0.02	0.10
1980/02	4544.91	6059.95	8614.30	509.81	0.02	0.11
1980/03	4724.54	6263.46	9442.07	680.57	0.02	0.12
1980/04	5270.47	6726.35	10091.92	498.58	0.02	0.12
1981/01	5957.58	8156.34	10741.06	530.40	0.02	0.13
1981/02	6333.60	8132.11	12009.58	970.44	0.02	0.14
1981/03	6435.75	6988.62	12067.27	541.95	0.03	0.15
1981/04	7141.08	7834.43	12534.18	1033.11	0.03	0.16
1982/01	8461.26	7018.25	11111.84	584.37	0.05	0.18
1982/02	9539.59	7708.70	10232.33	620.82	0.05	0.21
1982/03	10783.06	8038.15	8413.16	607.91	0.07	0.26
1982/04	12863.92	8206.95	7104.80	87.21	0.10	0.31
1983/01	15566.83	7546.83	5908.46	824.28	0.11	0.38
1983/02	17556.97	8115.75	6706.72	682.58	0.12	0.44
1983/03	19448.13	8270.07	7235.76	454.85	0.13	0.50
1983/04	22446.96	8995.99	7218.08	229.89	0.14	0.57
1984/01	26942.35	9768.88	7595.11	665.38	0.16	0.66
1984/02	29193.49	9502.06	8258.02	451.53	0.17	0.74
1984/03	32047.34	9288.21	8996.94	259.87	0.18	0.81
1984/04	35493.15	9271.35	8797.09	164.23	0.19	0.90
1985/01	43258.01	9245.22	9070.68	601.62	0.21	1.05
1985/02	47195.00	8379.06	8901.29	835.43	0.23	1.14
1985/03	51326.28	9110.34	8796.28	433.64	0.30	1.28
1985/04	58828.56	9124.28	8291.15	112.91	0.37	1.48
1986/01	65534.94	7517.33	7917.05	281.59	0.47	1.76
1986/02	76229.63	7342.09	8194.03	702.74	0.57	2.08
1986/03	84115.56	7024.96	7658.84	400.12	0.75	2.50
1986/04	103390.87	8043.47	7531.45	1016.25	0.92	3.04
1987/01	142013.13	8720.38	7341.77	352.52	1.12	3.76
1987/02	178761.17	9548.51	8025.85	632.51	1.35	4.72
1987/03	211755.84	9402.51	8871.71	497.00	1.57	5.88
1987/04	280832.44	9697.04	8890.15	1152.58	2.21	7.89
1988/01	374183.48	10339.17	9579.97	470.25	2.28	10.37
1988/02	414229.37	10749.38	10713.01	857.76	2.28	11.12
1988/03	410721.75	10343.14	11796.61	594.51	2.28	11.48
1988/04	452150.26	10664.10	12381.83	957.47	2.28	11.96
1989/01	503224.59	11696.86	12727.26	618.26	2.37	12.56
1989/02	545260.58	12346.97	13617.94	842.36	2.46	13.08
1989/03	541013.61	11708.15	13769.72	883.27	2.55	13.46
1989/04	590412.22	12351.40	13809.67	831.61	2.64	14.32
1990/01	657690.82	13121.84	15312.72	561.02	2.73	15.62
1990/02	712552.83	12468.74	14127.54	615.59	2.82	16.49
1990/03	732529.15	14203.82	15982.13	578.87	2.89	17.32
1990/04	836434.50	16276.49	18099.55	877.76	2.95	18.60
1991/01	876158.79	13390.64	15533.65	1649.37	2.98	19.69
1991/02	949987.30	14807.43	18659.34	1146.62	3.02	20.30

1991/03	922292.93	14606.24	18763.84	702.12	3.06	20.83
1991/04	1032321.50	15283.02	19777.22	1263.40	3.07	22.10
1992/01	1057357.07	14460.16	19611.50	1042.73	3.08	23.00
1992/02	1123373.63	15467.34	21365.15	1180.24	3.12	23.52
1992/03	1107118.94	15484.26	22339.25	1274.95	3.12	24.02
1992/04	1207896.17	16257.17	22791.51	894.88	3.12	24.74
1993/01	1221500.54	15628.31	21289.41	1163.68	3.10	25.40
1993/02	1250154.56	16951.87	22596.92	954.43	3.12	25.84
1993/03	1218290.07	16683.29	23349.43	550.09	3.12	26.30
1993/04	1334838.71	18488.60	23915.52	1720.60	3.11	26.72
1994/01	1355462.58	18061.83	24843.30	3152.00	3.36	27.21
1994/02	1424843.65	19406.26	26882.11	3283.37	3.39	27.61
1994/03	1384767.10	19460.75	27369.14	2813.92	3.40	28.06
1994/04	1528383.35	21442.92	28939.17	1723.21	5.33	28.61
1995/01	1629327.21	23017.42	24372.06	1982.82	6.82	32.76
1995/02	1794636.07	24056.12	23699.75	2913.61	6.31	38.02
1995/03	1806218.46	24413.15	24863.59	2254.72	6.42	40.26
1995/04	2131541.53	25542.59	25670.58	2375.16	7.64	43.47
1996/01	2283507.01	26626.22	26789.61	2027.67	7.55	47.10
1996/02	2453070.27	28416.71	28161.54	1779.94	7.61	50.12
1996/03	2488578.78	29104.36	29968.95	2004.38	7.54	52.34
1996/04	2894478.28	31168.84	32903.66	3373.46	7.85	55.51
1997/01	2948936.57	30415.41	30635.23	2109.24	7.89	58.62
1997/02	3138091.51	32658.92	33842.78	2594.52	7.96	60.32
1997/03	3091167.78	33334.00	35924.21	5594.98	7.82	62.15
1997/04	3538285.68	34909.83	38580.94	2530.82	8.08	64.24
1998/01	3659245.54	33703.57	36983.21	2615.01	8.52	67.57
1998/02	3756740.11	35475.58	38868.20	3469.92	9.04	69.56
1998/03	3777534.85	34206.83	38920.94	3277.21	10.11	72.05
1998/04	4199352.72	36762.49	41368.81	3302.12	9.87	76.19
1999/01	4324546.83	35424.32	39062.25	3503.55	9.52	79.90
1999/02	4528885.44	39268.39	42134.16	3373.22	9.49	81.66
1999/03	4529522.11	40820.89	43969.29	3013.88	9.36	83.46
1999/04	5018996.66	43396.94	47694.19	3932.30	9.51	85.58
2000/01	5304783.88	44325.80	49256.25	4456.80	9.23	87.98
2000/02	5453861.14	47462.68	50949.34	4724.75	9.95	89.34
2000/03	5432954.64	49489.04	53511.72	3016.08	9.41	90.84
2000/04	5799342.55	51598.23	57842.92	5754.94	9.57	93.25
2001/01	5815638.30	46942.07	51717.84	3513.91	9.54	94.30
2001/02	5796507.61	47806.18	51382.10	5109.28	9.06	95.21
2001/03	5660706.54	45788.07	49114.82	16265.95	9.53	96.42
2001/04	5974252.76	45630.10	51656.38	4638.91	9.14	97.35
2002/01	5905163.52	43355.87	47125.54	4976.56	9.03	98.69
2002/02	6319275.93	48427.04	51284.66	6208.09	10.00	99.92
2002/03	6168592.03	48089.56	51311.36	5990.55	10.17	101.19
2002/04	6676863.70	48267.35	52558.32	5879.92	10.31	102.90
2003/01	6732925.19	46649.38	48724.52	3714.96	10.77	104.26
2003/02	6904315.81	48248.10	49708.33	5103.03	10.48	104.19
2003/03	6668411.65	49812.67	51269.76	3642.30	10.93	105.28
2003/04	7275774.73	52070.78	54280.04	4222.18	11.24	107.00
2004/01	7374919.35	52019.31	53207.16	9022.02	11.15	108.67
2004/02	7615702.76	57702.29	56930.50	3878.89	11.41	108.74

2004/03	7543355.29	57547.78	58320.88	3004.05	11.41	110.60
2004/04	8321207.42	59234.75	63222.71	7038.26	11.26	112.55
2005/01	7976653.74	57097.88	59332.13	5841.84	11.29	113.44
2005/02	8324127.73	65213.56	65171.41	6366.43	10.84	113.45
2005/03	8173719.07	65117.38	65526.55	4955.86	10.85	114.48
2005/04	8990320.78	70494.88	72279.00	4665.45	10.78	116.30
2006/01	8819353.32	70323.01	72228.51	5345.84	10.95	118.94
2006/02	9493962.09	76460.27	76642.71	4835.43	11.40	122.02
2006/03	8944714.95	75481.21	75895.32	3082.31	11.02	123.17
2006/04	9372229.26	76495.72	78371.36	6067.81	10.88	122.71
2007/01	10631637.68	73032.00	78155.42	9977.32	11.08	125.13
2007/02	11107318.69	80513.18	82351.99	5673.52	10.87	126.26
2007/03	11292180.41	83378.01	83811.51	5573.90	10.92	127.46
2007/04	11793141.70	86959.64	87898.94	6303.36	10.87	129.79
2008/01	11579065.95	83059.74	85707.13	6066.60	10.70	132.87
2008/02	12338866.24	93015.28	95109.67	7140.01	10.28	136.28
2008/03	12402059.27	91110.37	95573.59	3804.37	10.52	137.65
2008/04	12122229.60	75540.90	82141.48	5470.18	13.49	135.65
2009/01	11162955.38	61539.50	64920.99	5326.49	14.63	139.28
2009/02	11400811.68	64523.23	64069.43	4649.47	13.53	140.31
Fuente: Elaboración propia en base a datos del INEGI y BANXICO.						

CAPITULO VIII.- LOS MODELOS CON CAMBIO ESTRUCTURAL

Para generar un modelo econométricamente la referencia básica es la teoría económica que el investigador desea expresar matemáticamente y verificar estadísticamente. Una vez que ésta está clara y que se dispone de la base de datos de las variables que se involucrarán en el modelo, éste se corre con el uso de la cibernética y algún software especializado, que en este caso es Eviews 5. Enseguida, al contarse con los resultados del modelo, si se observan cambios el investigador procede a verificar estadísticamente si ellos están sustentados o corresponden a la realidad económica que de maneras simple el modelo pretende representar apropiadamente; en otras palabras, procede a verificar estadísticamente si acepta dichos cambios; cuando así sucede: a).- Se dice que la teoría está bien descrita matemáticamente lo que significa que el modelo está bien especificado; si no, también estadísticamente puede ser que sugiera: b).- Que no está bien especificado, indicando por ejemplo, que c).- Sobran o que se omitieron variables explicativas de la realidad que el modelo pretende representar ; d).- Que el modelo no describa sucesos que ocurrieron inesperadamente como la crisis provocada por " el error de diciembre de 1994" o la crisis financiera de 2009, situación que orillará al investigador a buscar métodos más apropiados para modelar mejor las fluctuaciones económicas de la variable en estudio por el investigador.

Lo anterior se lleva a cabo aplicando todo el proceso que se describe a continuación:

VIII.1.-Especificación: Teoría económica.

Con base en estas referencias en que para configurar el modelo lo primero que se hace es establecer su objetivo, mismo que surge de las variables que se desean estudiar y que generalmente están estructuradas en torno a una teoría económica que, en otras palabras, constituye el marco teórico.

Para especificar el modelo es conveniente realizar una primera aproximación gráfica de la relación entre las variables, que nos indicarán el grado y la forma de relación existente entre ellas, la cual podría ser positiva (o negativa) lineal (o no lineal).

Una visión muy general (Carrascal, et al: 2000, 79) es asociar al análisis gráfico la matriz de correlaciones entre las variables, ya que cada uno de los coeficientes de correlación lineal entre cada par de variables (digamos Y_t, X_t) expresa su grado de asociación; mientras más se acerca su valor a –1 y +1 mayor será su relación y cuando se acerque a cero, ello indica; su escasa vinculación.

VIII.2.-Estimación: Método de mínimos cuadrados ordinarios: MCO.

Con ese análisis gráfico y numérico el analista está en condiciones de especificar el modelo de regresión lineal que le permitirá alcanzar el objetivo de su estudio, mismo que se construye generalmente con el método de MCO u otros, para obtener los estimadores o coeficientes de las variables explicativas de la variable explicada

VIII.2.1.-. Estimación de un modelo de regresión clásico.

VIII.2.1.1. Especificación y estimación de los parámetros del modelo con MCO.

Como se ha indicado, al expresar un modelo la relación analítica que existe entre variables con una estructura definida, esta debe de estar bien especificada para que represente adecuadamente la realidad que el investigador desea. Dicha relación matemáticamente se enuncia con las variables endógenas y exógenas y con los coeficientes de las mismas. En esta tesitura digamos concretamente que si el objetivo es explicar el comportamiento que observa una variable (Y) dependiente a partir de cambios en una (X) o varias independientes (X, Z, Q) también llamadas regresoras. Para ello se asume que existe una relación lineal entre ellas, tal que por ejemplo $Y_t=a+bX_t+cZ_t+dQ_t+e_t$

Donde

Y_t: Variable dependiente, endógena, explicada o regresada.

X_t, Z_t, Q_t; Variables independientes, exógenas, explicativas o regresoras.

E_t: Perturbaciones aleatorias

VIII.2.1.2. Análisis de las principales violaciones de los supuestos básicos de MCO

Cuando se estima el modelo con MCO se asume que se cumple una serie de supuestos o hipótesis clásicas o básicas, dentro de las que se vieron: la homocedasticidad, la ausencia de autocorrelación y de multicolinealidad, como condiciones básicas para que los estimadores (b, c, d) también llamadas pendientes o coeficientes de regresión es deseado que tengan las siguientes propiedades: sean eficientes, lineales, consistentes, insesgados y suficientes entre otras propiedades que deben tener para hacer confiable la estimación de Y con (Y_t), ya sea en forma determínistica o estocástica.

VIII.2.1.3. Contrastes de especificación y diagnóstico del modelo econométrico.

Contrastes de especificación: pruebas de hipótesis para verificar violaciones a los supuestos que fundamentan el método de cálculo MCO.

Una vez que se cuenta con el modelo, se estudia la posibilidad de que se hayan violado algunos de los principales supuestos básicos. Es por ello que se revisó si se habían violado los supuestos de homocedasticidad, independencia o ausencia de autocorrelación y de multicolinealidad. Cuando hubo violaciones, éstas se corrigieron para mejorar la calidad de los estimadores o la bondad de ajuste de la regresión lineal realizada con MCO.

VIII.2.1.4.- Errores de especificación en la selección de las variables.

Ahora se verán otras pruebas estadísticas relacionadas con la especificación con objeto de analizar su validez para sustentar la teoría económica que se desea exponer en el estudio.

Así, podemos decir que los errores específicos que se pueden cometer en la relación de las variables explicativas son dos:

1.- La omisión de variables relevantes.

2.- La inclusión de variables irrelevantes

Para verificar la presencia de estos dos errores se hace el siguiente análisis:

VIII.2.1.5. Análisis de la estabilidad estructural

Como señala Carrascal et al (ídem): "Una de las hipótesis que suponemos cumple el modelo de regresión especificado es que los coeficientes se mantiene constantes para todo el periodo muestral. Sin embargo, es posible que existan submuestras para las que el comportamiento del modelo, su estructura, sea diferente, siendo necesario contrastar esta posibilidad". Para ello aplicaremos:

VIII.2.1.5.1. Contraste y predicción de Chow

Para realizarlo la muestra total de datos se divide en varios grupos y se estima la ecuación cuya estabilidad se está evaluando para cada uno de ellos. Se establecen las dos siguientes hipótesis:

H_o: Hay un solo modelo para el conjunto de las observaciones: un modelo restringido que indica que hay estabilidad estructural.

H_a: Hay un modelo diferente para cada una de las submuestras en que se divide la muestra.

En este modelo sin restricciones los parámetros pueden cambiar de una submuestra a otra, es decir, no hay estabilidad estructural.

Observaciones: si no hay diferencias estadísticas significativas entre el modelo restringido y sin restringir no se rechaza Ho de estabilidad estructural del modelo.

De lo contrario se acepta Ha que indica que hay cambio estructural (en el caso de México este análisis es útil por ejemplo para ver si hubo un cambio estructural en 1995: crisis económica).

Para verificar que solo hay un cambio estructural se usa F, y para varios se usa el estadístico Chow, que proviene del estadístico de razón de verosimilitud; el cual se distribuye asintóticamente como una X^2 con grados de libertad igual al producto del número de cambios estructurales por el número de parámetros a estimar en el modelo restringido (Carrascal et al: 2000: 189). Así en el caso de F, ésta se compara con Fα y X^2 . En el pimer caso se dice que:

Si F empírica < Fα de tablas se acepta H_o

Si F empírica > Fα de tablas se rechaza H_o

VIII.2.1.5.2. Estimación recursiva.

María del Mar Zamora (2002, mariam.zamorauah.es) de la universidad de Alcalá de Henares, en la exposición del tema indica que una vez obtenidos los estimadores con MCO se debe proceder a *validar* el modelo para conocer la confiabilidad de sus resultados, sobre todo en su aplicación a la *predicción* de valores. Como se sabe, esta tarea se realiza con el instrumental de la prueba de hipótesis, para lo cual en general se usan los *residuos*, ahora llamados *recursivos* para con ellos calcular las estadísticas CUSUM, cuyos valores de ésta nos pueden señalar que los estimadores, por ejemplo, tienen heteroscedasticidad y autocorrelación. La pérdida de sus propiedades de homocedasticidad e independencia puede *corregirse* aplicando los residuos recurrentes, que se obtienen en forma recursiva o recurrente a partir de los parámetros β del modelo econométrico.

Esta estimación recursiva es similar a la MCO pero realizada de *manera recursiva aumentando el tamaño de la muestra gradualmente*.

Nosotros podemos agregar que ésta estimación es útil cuando los datos son temporales y se desconoce el momento en que se ha producido el cambio estructural en la serie temporal que estemos estudiando. Se hace la estimación secuencial del modelo especificado para distintos tamaños de muestras.

Aquí subyace la idea de que en este tipo de estimaciones si no hay cambio estructural las estimaciones de los parámetros se mantienen constantes al ir aumentando la muestra secuencialmente y los residuales no se desviarán ampliamente de cero.

Para verificar ésta hipótesis se *grafican los coeficientes y los residuos recursivos*. Si se observan grandes variaciones en la escala del eje de las ordenadas, al ir añadiendo nuevas observaciones a la muestra, ello indica que no hay estabilidad estructural y por consiguiente, que puede afectar las propiedades de los estimadores de las variables explicativas.

VIII.2.1.5.3. Coeficientes y Residuos recursivos: Estos son los errores de predicción de un periodo hacia adelante calculados en cada etapa de la estimación recursiva.

Se calculan y expresan gráficamente con Eviews que además calcula los estadísticos CUSUM y CUSUMQ, ambos construidos con los residuos recursivos.

Identificación: Cuando la gráfica muestra que uno o más residuos sobrepasan las bandas de confianza construidas en torno al CERO. Si eso sucede se dice que hay evidencia de que no hay estabilidad estructural.

Estadístico CUSUM

Viene dado por $w_t = \frac{\sum_{T=k+2}^{t} w_t}{S}$ donde t= k+2,…,T

Donde S es el error estándar de la regresión estimada con todas las observaciones disponibles. Bajo la Ho de estabilidad estructural Wt tiene u=0 por lo que sumas acumuladas que se alejan de u revelan que hay inestabilidad.

Estadístico CUSUMQ

A diferencia del CUSUM, el CUSUMQ se utiliza para verificar H_0: sumas acumuladas de los cuadrados de los residuos recursivos.

$$S_t = \frac{\sum_{T=k+2}^{t} w_t^2}{\sum_{T=k+2}^{T} w_t^2}$$ donde t= k+2,…,T

Bajo Ho de estabilidad de los parámetros, S_t tiene una esperanza matemática cuyo valor oscila entre 0 y 1.

El contraste se hace enfrentando los residuos St con t junto con sus bandas de confianza que se calculan con cierto grado de confianza (¶) y de error (α) o nivel de significación. Si la curva se sale de las bandas, hay ausencia de estabilidad estructural.

Solución: Comenta Carrascal et al(2000:200) que la solución a un problema de error de especificación por existencia de inestabilidad en el modelo debe resolverse incorporando el cambio estructural a la especificación del mismo. Para ello se debe indagar qué coeficientes de regresión son los afectados. Para saber donde ocurren los cambios, se sugiere obtener la estimación recursiva, y básicamente con la representación gráfica de los coeficientes recursivos. Posteriormente, la incorporación de variables ficticias, es una forma sencilla de averiguar que coeficientes son los afectados por un cambio estructural mediante la realización de contrastes de significación.

VIII.2.1.5.4. Errores de especificación en la forma funcional.

Estos errores se estudian utilizando el contraste RESET de Ramsey (1969). El contraste propone incorporar diversas potencias de Y_t que en realidad son potencias y productos cruzados de los regresores o variables explicativas.

Ho; Hay linealidad en el modelo

Ha: No hay linealidad en el modelo.

Para ello se usa F y X^2

Si se rechaza H_o la solución es determinar cuáles son las potencias y productos cruzados concretos de las variables explicativas que mejor recogen esa no linealidad.

A continuación se muestran ejemplos sobre modelos con cambio estructural usando Eviews.

VIII.3.- Ejercicios con Eviews

Ejercicio 1: Análisis sobre la especificación estructural del modelo

Sean los datos de las siguientes variables:

Tabla 8.3.1 Datos

Año	Y	X_1	X_2	X_3
1993	3	1	8	16
1994	2	2	15	17
1995	4	2.5	10	18
1996	5	3	9	15
1997	5	4	7	17
1998	7	5	6	20
1999	6	7	8	19
2000	8	8	4	21
2001	9	9	3	22
2002	12	15	1	23

ESPECIFICACION

I. Marco teórico: Teoría del consumo

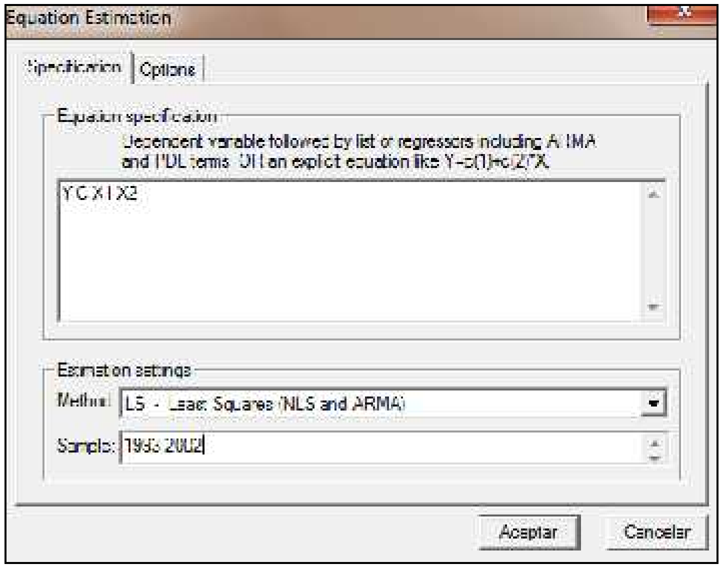
Con Y: consumo, X_1: Ingreso, X_2: Inflación, X_3: Inversión, con una serie de 10 años para cada una de las variables y haciendo

$Y=f(X_1, X_2)$

Verificar la teoría económica de que el consumo, Y, varia en razón directa del ingreso X_1, y en razón inversa de la inflación, X_2. Lo anterior significa, entre otras cosas, comprobar que el coeficiente del regresor o variable exógena, X_1, tiene signo positivo, en tanto que el coeficiente de la otra variable exógena, X_2, tiene signo negativo.

Así, vamos al programa Eviews una vez capturada la información, i.e., que ya se cuenta con el archivo, enseguida se coloca el cursor en Quick/ estimate equation/ escribimos Y C X_1 X_2/ Aceptar y aparece el siguiente cuadro:

Cuadro 8.3.1. Estimación de la ecuación.



En seguida aparecerán los resultados de la ecuación de regresión:

Cuadro 8.3.2. Resultados de la ecuación de regresión.

Dependent Variable: Y
Method: Least Squares
Sample: 2002 – 2011
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.800837	0.976443	5.940785	0.0006
X1	0.442193	0.076017	5.817014	0.0007
X2	-0.309751	0.081265	-3.811615	0.0066

R-squared	0.973305	Mean dependent var	6.100000
Adjusted R-squared	0.965678	S.D. dependent var	2.998148
S.E. of regression	0.555441	Akaike info criterion	1.905216
Sum squared resid	2.159601	Schwarz criterion	1.995991
Log likelihood	-6.526078	F-statistic	127.6122
Durbin-Watson stat	2.464587	Prob(F-statistic)	0.000003

Vemos que la ecuación de regresión o modelo uniecuacional es: $Y = 5.800837 + 0.442193X_1 - 0.309751X_2$ a la cual llamaremos "ecuación para hacer economía empírica, es decir para hacer análisis de estructura, para predecir o para evaluar políticas públicas"

II. Diagnóstico

1. La ecuación de regresión Y = 5.800837 + 0.442193*X_1 - 0.309750*X_2 en lo que se refiere a los signos de los coeficientes de X_1, X_2 cumplen con lo especificado por la teoría del consumo.
2. En lo que se refiere a las pruebas de significación estadística usando α=5% realizadas con t, para X_1, X_2en lo individual y F para la verificación conjunta de ambas, se detecta que tienen una probabilidad de casí cero, lo cual, corrobora que X_1 y X_2 explican satisfactoriamente a Y.
3. En ese sentido R^2= 0.973305 y la ajustada $\bar{R}^2$ = 0.965678 muestran que existe un alto grado de asociación de Y con X1, X2 de manera que no se necesita otra variable para determinar el comportamiento presente y futuro de Y.
4. Sin embargo, independientemente de verificar con este ejercicio si fueron o no violados algunos supuestos del modelo estimado con el método de mínimos cuadrados, digamos la homocedasticidad, la eficiencia, ausencia de autocorrelación o de multicolinealidad entre las variables exógenas, ahora se **harán otras pruebas adicionales con el propósito de cerciorarse que el modelo está bien especificado, i.e., que expresa satisfactoriamente la realidad que pretende representar**.

Dichas pruebas adicionales son:

III. Pruebas de especificación del modelo.

1.- Omisión de variables explicativas, digamos X_3.

Para verificar si omitimos algún regresor explicativo en el modelo aplicamos la $X^{2:}$ Chi cuadrada en su "razón de verosimilitud" como el estadístico F, cuya probabilidadindica si es necesario o no incorporar uno o más regresores explicativos y probar si su contribución al modelo es significativa estadísticamente. Como el lector puede observar, el uso de F se debe a que estamos interesados en indagar si debemos o no incorporar más variables explicativas al modelo de regresión, dicho en otra forma: nos importa probar estadísticamente la significación **conjunta** (efecto) de éstas variables en la variable dependiente o explicada. Así, planteamos las siguientes hipótesis con α≤5%: probabilidad de que los valores de los coeficientes de las variables explicativas estadísticamente sean iguales a cero.

H_o: uno o varios coeficientes de las variable(s) explicativa(s) **estadísticamente** no es (son) significativamente diferentes de cero. Si se verifica con F esta hipótesis nula (con una probabilidad mayor a 5%) decimos que dichos coeficientes no tienen un valor significativamente diferente de cero y por consiguiente, las variables explicativas a las que corresponden, no debe(n) incorporarse al modelo: en este caso X_3.

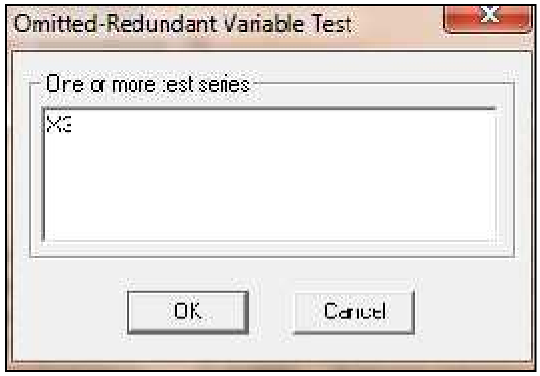
H_a: Todo lo contrario de H_o, es decir, si verificamos con F que los valores de los coeficientes es (son) significativo(s) estadísticamente, es decir, los valores de los coeficientes de las variables regresoras son distintos de cero, entonces debemos incorporarlo (s) al modelo porque también determina(n) el comportamiento de la variable endógena Y.

Para contrastar la contribución de las variables explicativas nuevas se requiere que éstas, también llamadas "omitidas", tengan el mismo tamaño de muestra que

Y, X_1, X_2; además, que el modelo especifique la ordenada al origen; C, y, Y, X_1, X_2.

Así, trabajando con el cuadro anterior de la ecuación de regresión antes descrita, partimos de su cuadro y nos colocamos en view/coefficient tests/omitted variables- likelihood ratio/y en la caja de "omitted – redundant variable test, escribimos X_3/ok y se obtiene el siguiente cuadro:

Cuadro 8.3.3. Prueba de Omisión de Variables



Como resultado de utilizar la variable X_3 tenemos el siguiente resultado:

Cuadro 8.3.4. Análisis de la Omisión de Variables.

Omitted Variables: X3				
F-statistic	0.193073	Probability		0.675754
Log likelihood ratio	0.316720	Probability		0.573585
Test Equation:				
Dependent Variable: Y				
Method: Least Squares				
Sample: 1993 2002				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.609288	2.903669	1.587402	0.1635
X1	0.407127	0.113580	3.584513	0.0116
X2	-0.303439	0.087583	-3.464580	0.0134
X3	0.071535	0.162801	0.439401	0.6758
R-squared	0.974138	Mean dependent var		6.100000
Adjusted R-squared	0.961206	S.D. dependent var		2.998148
S.E. of regression	0.590519	Akaike info criterion		2.073544
Sum squared resid	2.092274	Schwarz criterion		2.194578
Log likelihood	-6.367718	F-statistic		75.33212
Durbin-Watson stat	2.324024	Prob(F-statistic)		0.000037

Para **<u>evitar confusiones</u>** en este ejemplo es conveniente señalar que en este cuadro aparecen dos F´s. La diferencia entre ambas F's es la siguiente, la F= 0.193073 <u>está asociada a la prueba de omisión de variables</u> y la F= 75.33212 es la evaluación de la significación estadística conjunta de X_1,X_2 en la regresión Y= f(X_1,X_2) y, como se recuerda, manualmente se construye así: $F = \frac{R^2/k-1}{(1-R^2)/n-k}$

Comentarios

Como la probabilidad (0.675754) de la primera F y de la razón de verosimilitud (0.573585) son mayores que α=5%, se acepta Ho y se decide no incluir en el modelo X_3 como variable explicativa. Si se hubiera aceptado $H_a$se usaría la nueva ecuación de regresión $Y=f(X_1, X_2, X_3)$.

2. Variables explicativas redundantes, digamos X_2.

En Eviews la prueba “Redundant variables- likelihood ratio contrasta la significación estadística de uno o varios de los coeficientes de una o varias variables exógenas, con el fin de cerciorarse de que no sobran o de que son redundantes. Así con α=5% supóngase que deseamos verificar que:

H_o: X_2 es redundante

H_a: X_2 no es redundante

Como en el ejemplo anterior partimos del cuadro de la ecuación de regresión en que $Y=f(X_1, X_2)$, colocamos el cursor en view/coefficient tests/redundant variables-likelihood ratio/ escribimos X_2 en “omitted redundant variables test/ok y aparece el cuadro siguiente:

Cuadro 8.3.5. Prueba de Variables Redundantes.

Como resultado de utilizar la variable X_2 tenemos el siguiente resultado:

Cuadro 8.3.6. Variables Redundantes

Redundant Variables: X_2				
F-statistic	14.52841	Probability		0.006614
Log likelihood ratio	11.23463	Probability		0.000803
Test Equation:				
Dependent Variable: Y				
Method: Least Squares				
Sample: 1993 2002				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.263158	0.497606	4.548090	0.0019
X1	0.679087	0.071805	9.457435	0.0000
R-squared	0.917901	Mean dependent var		6.100000
Adjusted R-squared	0.907638	S.D. dependent var		2.998148
S.E. of regression	0.911169	Akaike info criterion		2.828679
Sum squared resid	6.641826	Schwarz criterion		2.889196
Log likelihood	-12.14339	F-statistic		89.44309
Durbin-Watson stat	2.526715	Prob(F-statistic)		0.000013

Vemos que las probabilidades de F y de X^2 en la razón de verosimilitud son menores que 5%, por lo que se acepta H_a y al no ser redundante el X_2, se conserva en el modelo como regresor explicativo de Y.

IV. Pruebas de estabilidad estructural

I. Contraste de Chow

Se usa cuando se piensa que dentro de la serie de años hay un año en que por razones extremas cambió la trayectoria de la variable endógena; en nuestro caso con 10 años suponga que en el año 6 ocurrió el desquiciamiento de la economía. Por el desequilibrio económico hay razones de peso para pensar que cambió el valor de Y. Para verificarlo, con α=5% se establece la hipótesis nula de estabilidad estructural.

H_o: Hay un solo modelo para todos los datos; ello se constata cuando la probabilidad de F es mayor que α=5%;

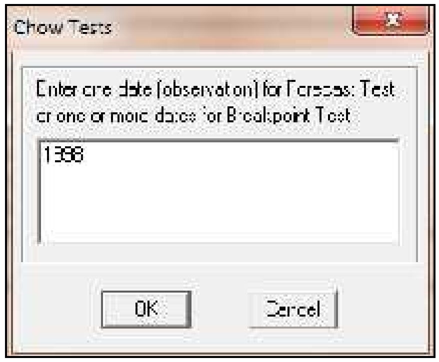
H_a: Cada subgrupo en que se dividen los datos tiene un comportamiento diferente: no hay un solo modelo hay dos; ello se constata cuando la probabilidad de F es menor que α=5%.

Luego si no hay diferencias estadísticas significativas entre el modelo restringido y el otro aceptamos Ho: de estabilidad estructural del modelo; en caso contrario, se acepta Ha y se dice que hubo un cambio estructural en el año 6 que fue 1998, en que todo el sistema económico se trastocó.

Para probar Ho se dividen en dos muestras los datos: la primera del año 1993 a y la segunda del año 1999 a 2002. Como siempre se parte del cuadro que muestra

$Y=f(X_1, X_2)$, de view/stability tests/chow break point test/ damos click y en la ventana que aparece se escribe/ok genera el cuadro:

Cuadro 8.3.7. Prueba Chow.



A continuación se mostrarán los resultados de la Prueba Chow para el año de 1998:

Cuadro 8.3.8. Resultado de la Prueba Chow.

Chow Breakpoint Test: 1998			
F-statistic	0.962950	Probability	0.491917
Log likelihood ratio	5.436099	Probability	0.142511

Al ver que la probabilidad de F y de la razón de verosimilitud es mayor que 5%, aceptamos H_0.

2. Contraste de predicción de Chow

Con α=5% se plantea:

H_0: Hay estabilidad estructural, en un modelo restringido a un solo comportamiento de todos los datos, cuyos residuos se comparan con los del período más largo de los dos en los que la serie ha sido dividida.

Se parte de la ecuación de regresión $Y=f(X_1, X_2)$ y se inicia el siguiente proceso: view/stability test/chow forecast tests/click y en "Chow test" se escribe el año en que pensamos a partir del cual pudo haber ocurrido un cambio estructural: 1998 / ok y aparece el cuadro:

Cuadro 8.3.9. Prueba de Pronóstico de Chow.

A continuación se mostrarán los resultados de la Prueba de Pronóstico de Chow para el año de 1998 a 2002:

Cuadro 8.3.10. Resultado de la Prueba de Pronóstico de Chow.

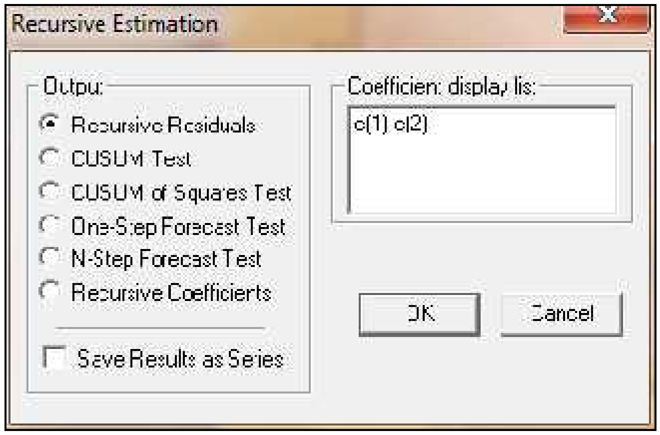
Chow Forecast Test: Forecast from 1998 to 2002				
F-statistic	0.630146	Probability		0.707345
Log likelihood ratio	9.459914	Probability		0.092067
Test Equation:				
Dependent Variable: Y				
Method: Least Squares				
Sample: 1993 1997				
Included observations: 5				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.376115	1.536202	2.848659	0.1043
X1	0.688921	0.305989	2.251456	0.1532
X2	-0.234532	0.109844	-2.135147	0.1663
R-squared	0.876682	Mean dependent var		3.800000
Adjusted R-squared	0.753364	S.D. dependent var		1.303840
S.E. of regression	0.647519	Akaike info criterion		2.252371
Sum squared resid	0.838561	Schwarz criterion		2.018034
Log likelihood	-2.630928	F-statistic		7.109128
Durbin-Watson stat	2.091172	Prob(F-statistic)		0.123318

De acuerdo con estos resultados se acepta H_0 ya que la probabilidad de los estadísticos es mayor que 5% y se concluye diciendo que hay estabilidad estructural, con esta muestra.

1. Estimación recursiva

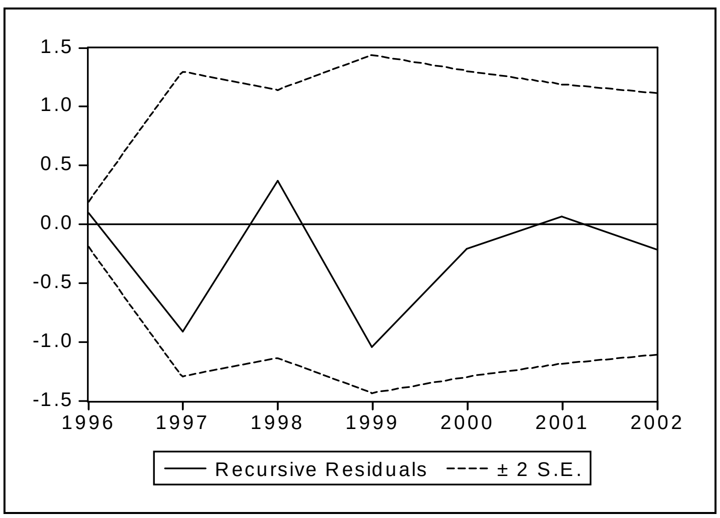
El método de los residuos recursivos o de los mínimos cuadrados recursivos es útil cuando se trabaja con datos temporales como estos y *se desconoce el año en que se produjo el cambio estructural*. Esta prueba consiste en la estimación secuencial del modelo especificado para distintos tamaños muéstrales. Si k=número de parámetros de las variables explicativas, entonces la primera muestra, su tamaño es k+1, que se usa para estimar el modelo: de 1993 a 1995 porque son dos parámetros. En las siguientes muestras se añade una a una todas las observaciones hasta agotar el total de la información (Carrascal; et al; 2000:193). De las sucesivas estimaciones del modelo con el resto de las muestras se generan “las series de los llamados coeficientes recursivos y residuos recursivos”. Si no hay cambio estructural: Ho. “Las estimaciones de los parámetros se mantendrán constantes al ir aumentando las muestras secuencialmente y los residuos no se desviarán amplia o significativamente de cero” Así, de View/stability tests/Recursive Estimate (OLS only)/ok y en la caja de diálogo aceptamos C(1) C(2), como puede apreciarse en el siguiente cuadro:

Cuadro 8.3.11. Estimación Recursiva.



Apareciendo la siguiente gráfica:

Grafico 8.3.1 Residuos



Claramente constatamos que hay estabilidad en el modelo en el período. Los residuos recursivos no se salen de las bandas construidas con dos errores estándar alrededor de Y ± 2S.E.

4. **Errores de especificación en la forma funcional/ RESET**

Estos errores se analizan con el contraste RESET elaborado por RAMSEY en 1969, *con el cual se verifica si se esta usando una forma funcional lineal incorrecta y cualquier error de omisión o la presencia de correlaciones entre las variables explicativas y la perturbación* (Carrascal et. al, 2000:203)

Con α=5% la probabilidad de rechazar una hipótesis cierta, se estable:

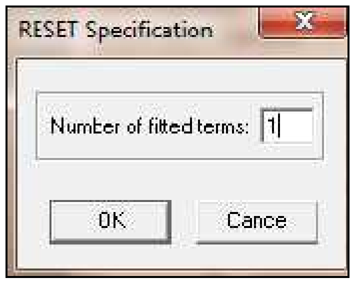
H_0: Hay linealidad en el modelo

H_a: No hay linealidad en el modelo

Dar View/Stability tests/Ramsey Reset Test/ clic y aparece el cuadro de diálogo “RESET Specification” y se escribe el *número de potencias* de la variable endógena ajustada a incluir empezando por el cuadrado; así si se indica se añadirá el cuadrado de dicha variable, si escribimos 2 se incluirá el cuadrado y el cubo, etc. Si se deja la celda en blanco el programa entiende que se añadirá la variable al cuadrado.

El resultado son las variables de F y X^2 de razón de verosimilitud junto con la ecuación estimada. Luego de View/ Stability tests/Ramsey RESET y se escribe en la celda en blanco 1 /ok y aparece el cuadro:

Cuadro 8.3.12. Especificación de la Prueba Ramsey RESET.



Se obtienen los siguientes resultados:

Cuadro 8.3.13. Resultados de la Prueba Ramsey.

Ramsey RESET Test:				
F-statistic	0.192822	Probability		0.675950
Log likelihood ratio	0.316314	Probability		0.573831
Test Equation:				
Dependent Variable: Y				
Method: Least Squares				
Sample: 1993 2002				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.911835	1.068461	5.533039	0.0015
X1	0.541787	0.240775	2.250178	0.0654
X2	-0.330648	0.098638	-3.352133	0.0154
FITTED^2	-0.011652	0.026536	-0.439115	0.6760
R-squared	0.974136	Mean dependent var		6.100000
Adjusted R-squared	0.961205	S.D. dependent var		2.998148
S.E. of regression	0.590531	Akaike info criterion		2.073584
Sum squared resid	2.092359	Schwarz criterion		2.194618
Log likelihood	-6.367921	F-statistic		75.32899
Durbin-Watson stat	2.643707	Prob(F-statistic)		0.000037

Al tener F y X^2 de razón de verosimilitud probabilidades mayores que 5% se acepta H_0 y se concluye que el modelo es lineal, que no hay omisión alguna y que no hay presencia de correlación entre las variables explicativas y la perturbación.

V. Normalidad entre las perturbaciones

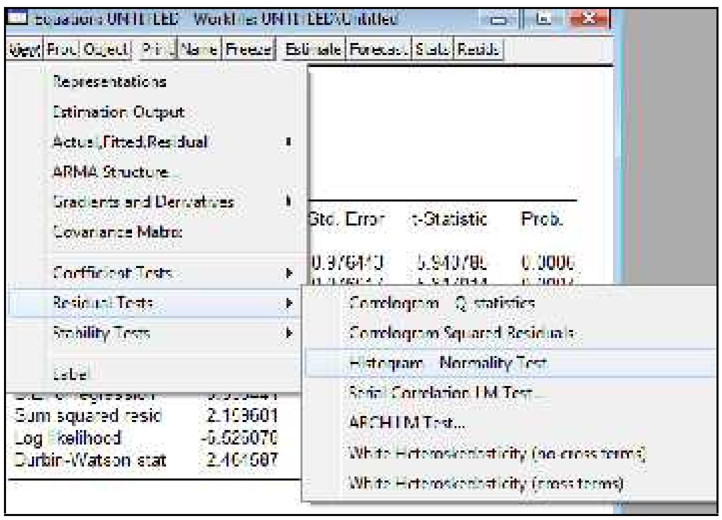
Este supuesto es básico para determinar el uso de otros métodos de estimación distintos al de MCO y para hacer inferencias a partir de la ecuación del modelo. Para ello es fundamental plantear con α=5% y verificar:

H_o: Hay normalidad en las perturbaciones cuando JB=0 y su probabilidad es mayor a α=5%;

H_a: No hay normalidad en las perturbaciones cuando JB ≠ 0 y su probabilidad es menor a α=5%.

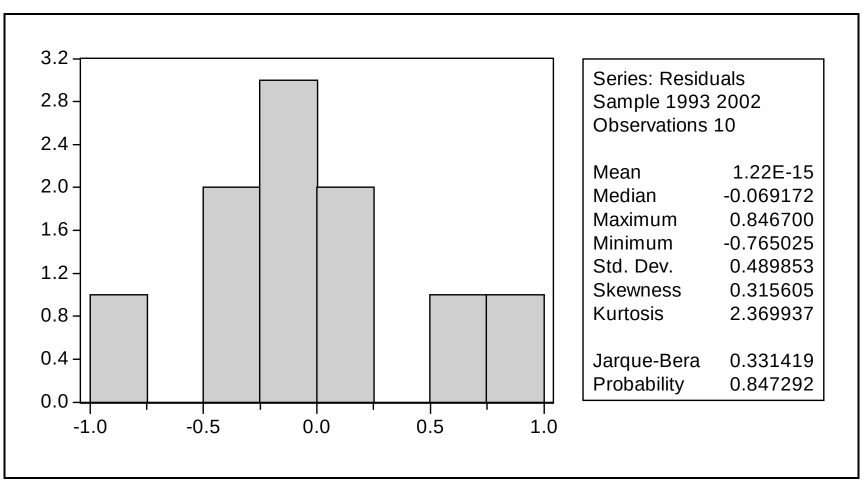
Como no son observables las perturbaciones se estudian con los residuos. Si se verifica Ho ello indica que la distribución empírica de los residuos debe ser similar a la de la distribución teórica de la normal. Para constatarlo vamos a: View/Residual tests/Histogram normality test/ click:

Cuadro 8.3.14. Prueba de Normalidad: Histograma.



Aparece la gráfica siguiente:

Grafico 8.3.2 Histograma.



Como la probabilidad de JB=0.847292 >α=5% y dado que JB se aproxima a cero se acepta Ho. Lo anterior se corrobora también con el valor de la kurtosis cercano a 3 y con el valor de la asimetría próximo a cero, indicando que se tiende a la normal.

Co este referente estadístico aceptable se concluye que la ecuación del modelo uniecuacional se puede usar para hacer econometría empírica, i.e., para hacer por ejemplo predicción, cuya metodología se expone a continuación.

VI. Predicción

1. Se Parte de File/open/workfile/clic/ aparece todo el archivo, ahí seleccionamos solo el archivo que llamamos "ecuación para predecir" /abrir/clic y aparece ese archivo.
2. Expandimos en 3 años el rango del archivo: "ecuación para predecir" con procs/change workfile range: start 1; and 3/ ok
3. Le indicamos a Eviews que ahora deseamos ampliar el tamaño de la muestra con: procs/Sample o sample en el workfile(archivo) en esa caja de diálogo ponemos 1993 a 2005/ok.
4. Damos en la ventana de la ecuación estimate: procs/make regresión group, aparecen Y, X_1, X_2 en blanco
5. Editamos pulsando edit/ escribir sus valores proyectados ver siguiente cuadro:

Tabla 8.3.2. Datos.

Obs	Y	X1	X2
1993	3	1	8
1994	2	2	15
1995	4	2.5	10
1996	5	3	9
1997	5	4	7
1998	7	5	6
1999	6	7	8
2000	8	8	4
2001	9	9	3
2002	12	15	1
2003			
2004			
2005			

Para llenar las celdas de X_1, X_2 hay 2 procedimientos:

1. Escribimos los valores proyectados de X_1 , X_2 (que son conocidos con antelación) como en el siguiente cuadro (llamado grupo 2) mismo que lo guardamos como grupo2
2. En la línea de comando escribir Data X_1 X_2 enter y aparece el cuadro del grupo 4 y llenamos los años con los datos correspondientes.

Grupo 2

Tabla 8.3.3 Datos

Obs	Y	X1	X2
1993	3	1	8
1994	2	2	15
1995	4	2.5	10
1996	5	3	9
1997	5	4	7
1998	7	5	6
1999	6	7	8
2000	8	8	4
2001	9	9	3
2002	12	15	1
2003		13	1
2004		14	2
2005		15	2

Grupo 4

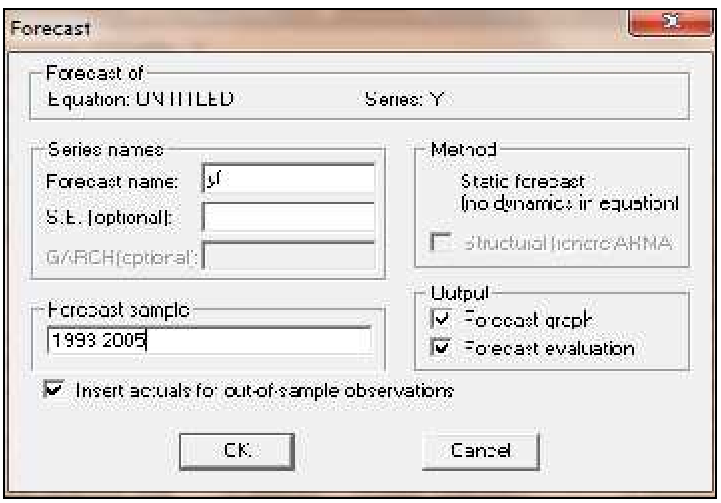
Tabla 8.3.4 Datos

Obs	X1	X2
1993	1	8
1994	2	15
1995	2.5	10
1996	3	9
1997	4	7

1998	5	6
1999	7	8
2000	8	4
2001	9	3
2002	15	1
2003	13	1
2004	14	2
2005	15	2

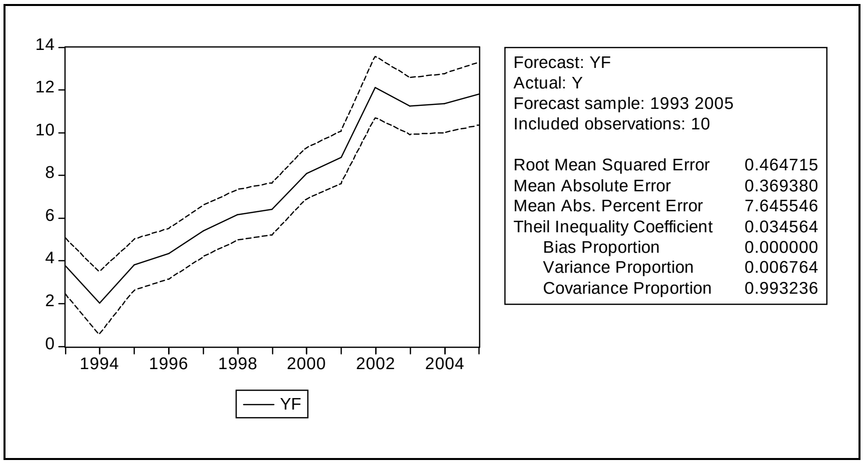
Regresamos a EQ01: que fue la primera ecuación en el workfile: “ecuación para predecir” y ahí voy a Forecast /aparece una pantalla que además de mostrar YF nos pide sample range for forecast”, ya se encuentra señalado el periodo del pronóstico 1993-2005, como se muestra a continuación:

Cuadro 8.3.15. Prueba de Pronóstico.



Los resultados de la gráfica son los siguientes:

Grafico 8.3.3 Predicción.



Cuyos resultados estadísticos del lado derecho sirven para evaluar la calidad de la predicción, mismos que se interpretarán párrafos abajo.

Ahora en la pantalla de workfile: ecuación para predecir, seleccione (sombreo) Y y YF para conocer el valor pronosticado de YF/ open group/clic y aparece el siguiente cuadro que llamo grup03

Tabla 8.3.5 Datos

Obs	Y	YF
1993	3	3.76502492522
1994	2	2.03896311067
1995	4	3.80881355932
1996	5	4.33966101695
1997	5	5.4013559322
1998	7	6.1533000997
1999	6	6.41818544367
2000	8	8.09938185444
2001	9	8.85132602193
2002	12	12.1239880359
2003		11.2396011964
2004		11.3720438684
2005		11.8142372881

Este cuadro también se obtendría así:

En la línea de comando escribimos “Show Y YF” / enter y aparece igual al caso anterior.

Evaluación de la Predicción: Puesto que ya conocemos YF con 13 datos, ahora es posible evaluar estadísticamente dicha predicción analizando en la figura anterior la gráfica y las estadísticas del cuadro. En el caso de la grafica se observa primero una caída y luego un repunte; ello se debe a que X_1, X_2 en el período de predicción, primero caen sus valores y luego aumentan.

Con respecto a las estadísticas, dado que casi todas (con la excepción de una: mean abs, percent error) tienen valores menores a uno, se dice que es buena estimación de la predicción. Esto también se constata con el pequeño valor de Theil, que se aproxima a cero.

Por consiguiente, para predecir, resultó adecuada la ecuación de regresión $Y=f(X_1, X_2)$

Ejemplo 2:

A.-Enfoque clásico

De acuerdo con la Teoría Económica se especifica que la Renta Neta Disponible (Carrascal; et al; 2000:167) de una Familia :**RNDFAM**, depende o esta relacionada con los Impuestos Directos pagados por las Familias :**TDFAM**, y el Ahorro Neto de las Familias, **SNFAM.**

Para verificar esta teoría económica se obtuvieron sus series de datos correspondientes al periodo 1964-1998, las cuales aparecen enseguida:

Tabla 8.3.6 Datos

Obs	RNDFAM	TDFAM	SNFAM
1964	908484.7	21493.29	79715.31
1965	1109146.	23669.82	138476.9
1966	1288729.	28107.12	175705.3
1967	1406532.	31627.71	157508.7
1968	1556219.	33935.14	164527.4
1969	1715110.	37408.60	181566.9
1970	1918993.	45636.83	216098.3
1971	2188887.	56869.02	259892.9
1972	2553119.	67278.84	305740.9
1973	3087734.	88962.40	390348.0

1974	3814323.	113104.3	476607.0
1975	4470742.	149828.1	544914.4
1976	5374717.	196773.1	549558.9
1977	6719341.	274270.0	659282.7
1978	8223260.	421734.4	939086.6
1979	9537096.	556372.6	942408.2
1980	10973779	799621.0	903711.0
1981	12501555	925866.0	1045899.
1982	14424134	981530.0	1282386.
1983	16136560	1313931.	1328501.
1984	17759873	1604035.	1415125.
1985	19770838	1834309.	1680170.
1986	22191064	1960032.	1923446.
1987	24097082	2741669.	1382455.
1988	26685069	3151992.	1878784.
1989	29773332	3860443.	1623172.
1990	33708161	4297860.	2611061.
1991	37361755	4962487.	3357835.
1992	39989229	5751092.	2858946.
1993	42733507	5741111.	4387816.
1994	44049168	6031643.	3368167.
1995	47756339	6333899.	4529397.
1996	50104688	6645635.	4569628.
1997	52104343	7139426.	3883061.
1998	54761946	7376184.	3706889.

Una ves definidos e identificadas las variables a relacionar mediante un modelo uniecuacional, digamos para propósitos ilustrativos: **RNDFM = F (TDFAM),** se debe de empezar por verificar gráficamente la posible relación entre ellas mediante la elaboración de su grafica correspondiente (diagrama de dispersión) para ver si su asociación es o no lineal. Enseguida, estimar la ecuación de regresión cuyos coeficientes de las variables regresoras al igual que los coeficientes de correlación parciales (simples) y múltiple indicarán el grado de vinculación entre la variable dependiente, RNDFAM, y la independiente TDFAM.

Resultados

Así, en el siguiente cuadro se verifica la Teoría Económica sobre la relación y asociación entre RNDFAM y TDFAM, cuya interpretación de los valores de R^2, R^2 (ajustada),coeficientes de C y TDFAM, al igual que las t´s y F permite concluir que el modelo representa adecuadamente (*con la excepción del signo del coeficiente de TDFAM*, cuya corrección se hará mas adelante). <u>En este contexto ahora profundizaremos el análisis para tener la certeza de que su ecuación de regresión realmente esta "limpia" de perturbaciones</u> y puede usarse para hacer los análisis de estructura, dinamismo, predicción, planeación y evaluación de políticas económicas.

Cuadro 8.3.16. Asociación entre RNDFAM y TDFAM

Dependent Variable: RNDFAM
Method: Least Squares
Sample: 1964 1998
Included observations: 35

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3693463.	492902.8	7.493288	0.0000
TDFAM	6.924402	0.149168	46.42027	0.0000

R-squared	0.984917	Mean dependent var	18650139
Adjusted R-squared	0.984460	S.D. dependent var	17702212
S.E. of regression	2206780.	Akaike info criterion	32.10741
Sum squared resid	1.61E+14	Schwarz criterion	32.19629
Log likelihood	-559.8797	F-statistic	2154.841
Durbin-Watson stat	0.296556	Prob(F-statistic)	0.000000

Este nuevo análisis se inicia aplicando las pruebas de especificación del modelo; posteriormente serán otras hasta asegurarnos que la ecuación de regresión garantiza resultados confiables para hacer la econometría empírica descrita al final del párrafo anterior.
Por consiguiente, iniciemos el siguiente proceso de Verificación.

Pruebas de especificación del modelo.

1.- Omisión de Variables Explicativas.

Para verificar si omitimos algún regresor o variables explicativas en el modelo

RNDFAM = 3693463+6.924402TDFAM, aplicamos la razón de verosimilitud X^2 que permite incorporar una o varias variables explicativas y probar si su contribución al modelo es significativa estadísticamente. Así tenemos que :

H_0: Una o varias variables no es (son) significativa (s) estadísticamente. No debe (n) incorporarse al modelo.

H_a: Todo lo contrario de Ho, es decir, que si se verifica que es (son) significativa (s) estadísticamente, entonces debemos incorporarla (s) al modelo por que también determina (n) el comportamiento de la variable explicada o endógena.

Para contrastar la contribución de la variables explicativas nuevas a la ecuación de regresión, se requiere que estas *también llamadas "omitidas"*, tengan el mismo tamaño de muestra que RNDFAM y TDFAM; además de que el modelo se haga especificado describiendo las variables y ordenada al origen.

Con estas referencias, digamos que deseamos contrastar si la nueva variable explicativa SNFAM (ahorro neto familiar) fue omitida y cuyos datos aparecen en la tercer columna del cuadro.

En Eviews una vez que hemos obtenido el cuadro de la ecuación arriba mencionada, vamos a view/ coefficient Test/ Omitted Variables-likelihood Ratio/ en caja de "Omitted-Redundant variable test , escribimos SNFAM/ OK y obtenemos.

Cuadro 8.3.17. Variables redundantes

Redundant Variables: SNFAM				
F-statistic	11.74974	Probability		0.001692
Log likelihood ratio	10.94624	Probability		0.000938
Test Equation:				
Dependent Variable: RNDFAM				
Method: Least Squares				
Date: 05/02/06 Time: 10:20				
Sample: 1964 1998				
Included observations: 35				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3693463.	492902.8	7.493288	0.0000
TDFAM	6.924402	0.149168	46.42027	0.0000
R-squared	0.984917	Mean dependent var		18650139
Adjusted R-squared	0.984460	S.D. dependent var		17702212
S.E. of regression	2206780.	Akaike info criterion		32.10741
Sum squared resid	1.61E+14	Schwarz criterion		32.19629
Log likelihood	-559.8797	F-statistic		2154.841
Durbin-Watson stat	0.296556	Prob(F-statistic)		0.000000

Con α= 5% = probabilidad de cometer error tipo I si Ho es cierta , se observa que la probabilidad de F (0.001692) y con Log Likelihoold Ratio (0.00938) es menor que α, luego entonces se rechaza Ho y se acepta H_a: SNFAM es significativa estadísticamente y <u>no debe omitirse sino incluirse en el modelo</u>.

1. **Variables Explicativas redundantes.**

En Eviews la prueba “Redundant Variables-likelihood Ratio contrasta la significación estadística de una o varias variables exógenas, con el fin de cerciorarse de que no sobran o de que son redundantes. Así, supóngase que:

Cuadro 8.3.18. Variables explicativas redundantes

Dependent Variable: RNDFAM				
Method: Least Squares				
Sample: 1964 1998				
Included observations: 35				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2615779.	531132.8	4.924906	0.0000
TDFAM	5.338268	0.480521	11.10932	0.0000
SNFAM	2.923527	0.852890	3.427790	0.0017
R-squared	0.988968	Mean dependent var		18650139
Adjusted R-squared	0.988278	S.D. dependent var		17702212
S.E. of regression	1916586.	Akaike info criterion		31.85181
Sum squared resid	1.18E+14	Schwarz criterion		31.98512
Log likelihood	-554.4066	F-statistic		1434.265
Durbin-Watson stat	0.248074	Prob(F-statistic)		0.000000

H_0: SNFAM es redundante

H_a: SNFAM no es redundante.

Para verificar la H_0 se parte de la ecuación de regresión anterior:

RNDFAM =2615779+5.338268TDFAM+2.923527 SNFAM vamos a view/ coefficient test /Redundant Variables – likelihood ratio/ se escribe SNFAM, en “Omitted-Redundant Variable test/ OK y aparece el cuadro del dialogo con la ecuación originalmente estimada con

MCO: RNDFAM =3693463+6.924402 TDFAM, acompañada de los valores de F y la razón de verosimilitud, con sus respectivas probabilidades, estos últimos sirven para probar la Ho con α=5%.

Como las probabilidades de F (0.001692) y de la razón de verosimilitud (0.000938) son menores que 5%, se rechaza H_0: es decir, se rechaza que SNFAM no sea relevante para explicar la variable dependiente RNDFAM. Decimos que SNFAM, si tiene poder explicativo de RNDFAM y debe permanecer en el modelo por que no es redundante.

Para ilustrar con este ejercicio la aplicación de otras pruebas para obtener la ecuación adecuada de la realidad que se desea expresar, se continuará trabajando con el caso RNDFAM =C+TDFAM, independientemente de que al final se resuman todos los resultados, en particular, los que muestran

problemas de identificación del modelo, mismos que darán la pauta para hacer las adecuaciones pertinentes como digamos, la inclusión de la variable SNFAM a la ecuación de regresión y la corrección del signo del coeficiente de la variable TDFAM. Así:

Pruebas de estabilidad estructural.

I. Contraste de Chow.

Para realizarla antes se establece la hipótesis nula de estabilidad estructural;

H_o: hay un solo modelo para todos los datos /modelo restringido).

H_a: cada subgrupo en que se dividen los datos tiene un comportamiento diferente.

Luego si no hay diferencias estadísticas significativas entre el modelo restringido y el otro se acepta Ho de estabilidad del modelo; en caso contrario, aceptamos Ha y decimos que hay cambio estructural. Para ello fijamos α= 5%.

Para ello dividiremos en 2 muestras nuestras observaciones: la primera de 1964 a 1987, por que en este ultimo año el índice inflacionario alcanzo su mas alto valor: 159.7 % situación que sin duda afecto el ingreso familiar, de ahí que la segunda muestra sea de 1988 a 1998.

Usando el programa Eviews, en la ecuación de regresión vamos a view, de ahí a "stability test" y enseguida a "Chow break point test", hacemos clic y en la ventana de Chow tests escribimos 1987 por ser el año que divide los dos subgrupos o muestras de datos y damos OK, de manera que aparecen los estadísticos para contrastar el cambio estructural a partir de 1987 el F y el Log Likelihood Ratio (razón de verosimilitud) con sus probabilidades.

LS RNDFAM C TDFAM

Cuadro 8.3.19. Prueba Chow.

Chow Breakpoint Test: 1987			
F-statistic	41.98661	Probability	0.000000
Log likelihood ratio	45.87492	Probability	0.000000

Al ser 0 la probabilidad de F se rechaza H_0 ya que para aceptarla era necesario que dicha probabilidad fuera mayor que 5%, se dice que en el modelo estimado se produce un cambio estructural en el año de 1987.

II. Contraste de Predicción de Chow.

Ho: Hay estabilidad estructural, es un modelo restringido a un solo comportamiento de todos los datos, cuyos residuos se comparan con los del *periodo mas largo* de los dos en que la serie ha sido dividido.

Como en el caso anterior, en el cuadro de dialogo de la ecuación de regresión vamos a view, luego a stability tests y enseguida a Chow forecast test/ clic y en "Chow test" se escribe el año en que pensamos a partir del cual pudo haber ocurrido un cambio estructural: 1987, decimos OK y aparece:

Cuadro 8.3.20. Prueba Chow.

Chow Forecast Test: Forecast from 1987 to 1998				
F-statistic	8.367933	Probability		0.000015
Log likelihood ratio	61.41428	Probability		0.000000
Test Equation:				
Dependent Variable: RNDFAM				
Method: Least Squares				
Sample: 1964 1986				
Included observations: 23				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2104179.	309107.7	6.807266	0.0000
TDFAM	10.48158	0.387631	27.04011	0.0000
R-squared	0.972081	Mean dependent var		7375228.
Adjusted R-squared	0.970751	S.D. dependent var		6727062.
S.E. of regression	1150481.	Akaike info criterion		30.83220
Sum squared resid	2.78E+13	Schwarz criterion		30.93094
Log likelihood	-352.5703	F-statistic		731.1674
Durbin-Watson stat	0.382958	Prob(F-statistic)		0.000000

Con estos resultados se vuelve a rechazar H_0 ya que la probabilidad no es mayor a 5% y se concluye diciendo que no hay estabilidad estructural en el modelo.

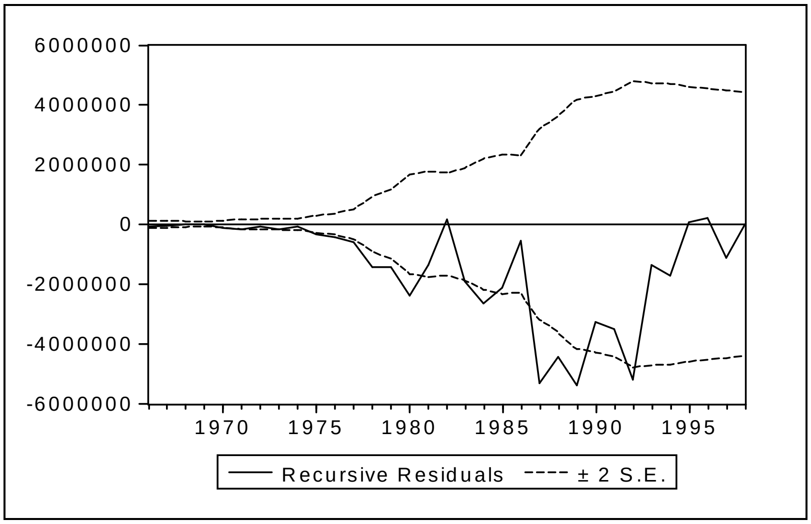
III. Estimación Recursiva.

Como antes se indicó, esta prueba es útil cuando se trabaja con datos t*emporales* como estos y se *desconoce* el año o momento en que se produjo el cambio estructural.

Esta estimación consiste (Carrascal; et al;2000:193)en la estimación secuencial del modelo especificado para distintos tamaños muéstrales. Por su importancia diremos que "si el numero de parámetros del modelos es k+1", ese será el tamaño de la primera muestra que usaremos para estimarlo; en las siguientes se añade una a una todas las observaciones hasta agotar el total de la información. De las sucesivos estimaciones del modelo con el resto de las muestras se generan "las series de los llamados coeficientes recursivos y residuos recursivos". Si no hay cambio estructural: H_0. Las estimaciones de los parámetros se mantendrán constantes al ir aumentando la muestra secuencialmente y los residuos no se desviarán ampliamente de cero."

Así en view vamos a Stability test /Recursive Estimate(OLS only)/Ok y en la caja de dialogo por default aceptamos C(1) C(2)/OK y aparece la siguiente grafica, que claramente muestra inestabilidad el modelo en el periodo.

Grafico 8.3.4 Estimación Recursiva.



Se observa que al ir incorporando datos a la muestra conforme se realiza la estimación, gradualmente se detecta el comportamiento de los estimadores dentro de las bandas de confianza construidas a $\pm$ dos veces su σ, es decir, los coeficientes varían mucho con el cambio del tamaño de la muestra y por consiguiente, no existe estabilidad estructural.

IV. Errores de Especificación en la forma funcional.

Como se indicó, estos se analizan con el contraste RESET elaborado por Ramsey en 1969, el cual permite identificar si se esta usando una forma lineal incorrecta y cualquier error de omisión o la presencia de correlaciones entre las variables explicativas y la perturbación (Carrascal; et al; 2000:203).

De view nos vamos a Stability Test/ Ramsey RESET TEST/ clic y aparece el cuadro de diálogos "RESET Specification"y escribimos el numero de potencias de la variable endógena ajustada a incluir empezando por el cuadrado; así, si indicamos 1 se añadirá el cuadrado de dicha variable; si ponemos 2 se incluirá el cuadrado y el cubo, etc. Si dejamos en blanco la celda el programa entiende que se añade la variable al cuadrado.

El resultado son las variables de F y X^2 de razón de verosimilitud junto con la ecuación estimada. Así en Eviews hacemos view/ Stability Test/ Ramsey RESET y escribimos en la celda en blanco 1, OK y aparece el cuadro:

Cuadro 8.3.21. Prueba Ramsey

Ramsey RESET Test:				
F-statistic	12.97673	Probability		0.001055
Log likelihood ratio	11.91433	Probability		0.000557
Test Equation:				
Dependent Variable: RNDFAM				
Method: Least Squares				
Sample: 1964 1998				
Included observations: 35				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2912145.	474658.0	6.135248	0.0000
TDFAM	9.024595	0.596848	15.12043	0.0000
FITTED^2	-5.80E-09	1.61E-09	-3.602323	0.0011
R-squared	0.989269	Mean dependent var		18650139
Adjusted R-squared	0.988598	S.D. dependent var		17702212
S.E. of regression	1890262.	Akaike info criterion		31.82415
Sum squared resid	1.14E+14	Schwarz criterion		31.95746
Log likelihood	-553.9226	F-statistic		1474.939
Durbin-Watson stat	0.407414	Prob(F-statistic)		0.000000

Si H_o: Hay linealidad en el modelo

H_a: No hay linealidad en el modelo

Con α= 5 % = probabilidad de rechazar una hipótesis cierta, decimos que al tener F y X^2 probabilidades pequeñas menores a 5 % ello indica que debamos rechazar H_0, y concluir que no hay linealidad (aceptar H_a) e ir pensando en cuales son las potencias y productos cruzados (que se verán mas adelante)concretos de las variables explicativas que mejor recogen esa no linealidad.(Carrascal 2000:205).

V. Normalidad de las Perturbaciones.

Este supuesto es básico para el uso de otros métodos de estimación distintos al de MCO y para hacer inferencias a partir del modelo. Por ello es fundamental plantear y verificar las,

H_0: Hay normalidad en las perturbaciones JB = 0

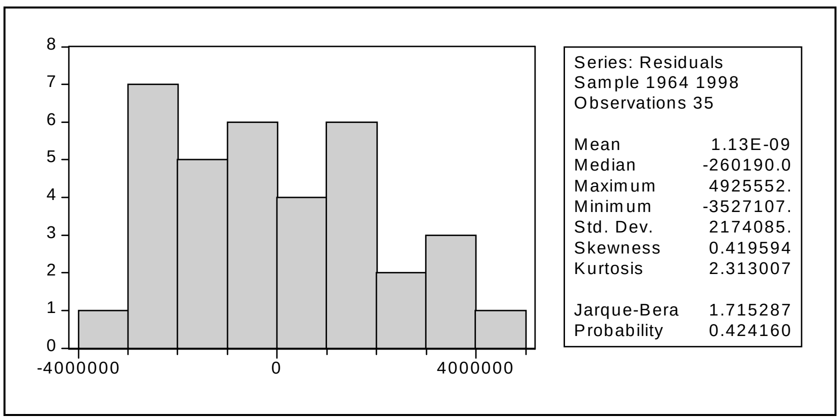
H_a: No hay normalidad en las perturbaciones JB $\neq$ 0.

Como no son observables las perturbaciones al estudio de su normalidad se realiza con los residuos. Si se verifica Ho la distribución <u>empírica</u> de los residuos debe ser <u>similar</u> a la de la distribución normal.

El estadístico Jarque Bera contrasta la normalidad de una variable, es decir, permite encontrar valores similares a los momentos poblacionales cuando se calculan los momentos muéstrales de los residuos (en una serie los momentos impares de una variable normal son cero y también su coeficiente de asimetría y su kurtosis próxima a 3). Eviews da el histograma de los residuos y el valor de Jarque Bera.

Hacemos: view/ Residual test/ Histogram normality test/ clic y ambos aparecen. Vemos que el histograma no es como la normal; el coeficiente de asimetría (Skewness) no es cero, pero la Kurtosis es cercana a 3 y quizás por ello el estadístico de Jarque Bera acepta Ho por que su probabilidad (0.42) de rechazarla siendo cierta es mayor que 5% =α.

Grafico 8.3.5 Histograma



En otras palabras como JB = 1.715287 tiende a cero y puesto que es menor que X^2 = 18.0 y dado que su probabilidad es mayor que 5% aceptamos Ho y decimos que <u>el modelo respeta el supuesto de normatividad.</u> Indudablemente que a medida que JB tiende a cero, la curva de la ecuación de regresión tenderá a la normal. Al ser JB =1.715287 con una asimetría= 0.419594(positiva) explica que la distribución de datos está ligeramente cargada a la derecha y como Kurtosis = 2.313007, que la curva es “achatada” o platocúrtica. De ahí que la media si tiende a cero. (0.000000000113) aunque su desviación estándar (2174085) es alta, y cuya varianza debe ser constante, indica que los dato tienen cierta variabilidad, motivo por el cual el sesgo o asimetría no sea cero, como tampoco la kurtosis es igual a 3.

Concluimos señalando que pueden aplicarse otros métodos distintos del MCO, con las reservas del caso. <u>También vemos que el modelo no pasa la mayoría de las pruebas aplicadas, por lo que no sirve para los objetivos establecidos.</u>

ANTES, HAGAMOS EL INTENTO DE RESOLVER O ENCONTRAR UNA SOLUCIÓN AL PROBLEMA DE SIGNO.

Para resolver el problema de los signos que fija la Teoría Económica, los autores en el mejor de los casos dicen cómo resolverlo pero generalmente no hacen ejercicios para ilustrar sus sugerencias metodólogicas. En mi libro de "Introducción a la Econometría" introduzco el método de la "deflactación de los valores de las variables" con el cual resuelvo el problema. Aquí diremos adicionalmente que hay opciones como:

(1) Modificar la teoría: ahora digamos SNFAM=f(RNDFAM,TDFAM)
(2) Aumentar la muestra del fenómeno en estudio.
(3) Transformar las variables en logaritmos.
(4) En arch, ver ejemplos: logaritmos y formas funcionales lineales

Así veamos con la opción (1):

Cuadro 8.3.22. Análisis de la Ecuación de la Regresión.

Dependent Variable: SNFAM				
Method: Least Squares				
Sample: 1964 1998				
Included observations: 35				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	29327.91	124724.6	0.235141	0.8156
RNDFAM	0.091864	0.026800	3.427790	0.0017
TDFAM	-0.093563	0.186988	-0.500366	0.6202
R-squared	0.946834	Mean dependent var		1540511.
Adjusted R-squared	0.943511	S.D. dependent var		1429436.
S.E. of regression	339740.7	Akaike info criterion		28.39157
Sum squared resid	3.69E+12	Schwarz criterion		28.52489
Log likelihood	-493.8525	F-statistic		284.9423
Durbin-Watson stat	1.726740	Prob(F-statistic)		0.000000

En este caso, se observa que los signos son los esperados por la teoría, pero no así la significancia estadística del coeficiente de la variable TDFAM y esto con un R^2 de 94.68% es un claro síntoma de multicolinealidad en el modelo, además de que el error estándar de la estimación también es muy alto y manifiesta que existe una gran variabilidad conjunta.

Trabajando (3):
Ahora si la estimación se realiza usando logaritmos obtenemos

Cuadro 8.3.23. Análisis de la Ecuación de Regresión con Logaritmos.

Dependent Variable: LSNFAM				
Method: Least Squares				
Sample: 1964 1998				
Included observations: 35				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-9.899852	2.081677	-4.755711	0.0000
LRNDFAM	2.149923	0.277244	7.754637	0.0000
LTDFAM	-0.827102	0.179860	-4.598586	0.0001
R-squared	0.989286	Mean dependent var		13.69985
Adjusted R-squared	0.988616	S.D. dependent var		1.180299
S.E. of regression	0.125931	Akaike info criterion		-1.224346
Sum squared resid	0.507477	Schwarz criterion		-1.091031
Log likelihood	24.42606	F-statistic		1477.367
Durbin-Watson stat	1.530522	Prob(F-statistic)		0.000000

I. Ahora se observa que el signo del coeficiente de la variable LTDFAM es el esperado, pero el signo de la ordenada al origen no lo es, lo cual no es trascendente por que indica la parte autónoma u original que la teoría económica no considera necesaria conocer o bien, dicho en otras palabras, su punto de partida geométricamente es la pendiente;ahora bien, si la teoría hubiera establecido al ahorro como indispensable, entonces intentaríamos resolver el problema del signo con cualesquiera de los cuatro procedimientos señalados en el punto V.-Normalidad de las Perturbaciones, página 44.

Por otra parte, la significación estadística de los coeficientes de las variables independientes es buena y la prueba global también es significativa: las varianzas son pequeñas de los coeficientes y de la regresión, el estadístico DW se encuentra en la zona de aceptación de la H_0, por lo que con las reservas del caso antes señaladas, el modelo en general esta correctamente especificado.
Aquí ya no se juzgo necesario hacer la predicción porque ésta no es el tema principal del ejercicio, independientemente de que ya existen muchos ejercicios previos en los que se ilustra cómo realizarla.

Ejemplo 3:

I. Marco teórico.

Haciendo uso de la base de datos de la economía mexicana consideramos las siguientes variables: las Exportaciones de bienes y servicios (X) en millones de dólares, el Producto Interno Bruto (PIB) expresado en miles de pesos a precios constantes de 2003, la Formación Bruta de Capital Fijo (FBKF) en miles de pesos a precios de 2003 y el Índice Nacional de Precios al Consumidor (INPC) representado con porcentaje. Para llevar a cabo la regresión se utilizarán datos a partir del primer trimestre de 1985 hasta el cuarto trimestre de 2010, la ecuación se determina de la siguiente manera:

$$X = f(PIB, FBKF)$$

Dentro de la teoría económica las Exportaciones varían en relación directa tanto en el Producto Interno Bruto como en la Formación Bruta de Capital Fijo. Lo anterior significa, comprobar que el coeficiente del regresor o de las variables exógenas PIB y FBKF tiene signo positivo.

Para hacer la regresión consideramos los pasos que anteriormente se han mencionado para establecer la regresión (véase cuadro 1), en este caso paras las variables que señalamos en los párrafos anteriores.

Cuadro 8.3.24. Resultados de la ecuación de regresión.

Dependent Variable: X
Method: Least Squares
Date: 09/12/12 Time: 20:20
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-70870.44	2720.019	-26.05512	0.0000
PIB	1.42E-05	1.01E-06	13.98035	0.0000
FBKF	1.11E-05	3.64E-06	3.063378	0.0027

R-squared	0.960721	Mean dependent var	33392.56
Adjusted R-squared	0.960072	S.D. dependent var	26035.35
S.E. of regression	5202.409	Akaike info criterion	19.97553
Sum squared resid	3.27E+09	Schwarz criterion	20.04376
Log likelihood	-1235.483	F-statistic	1479.756
Durbin-Watson stat	0.433006	Prob(F-statistic)	0.000000

II. Diagnóstico.

a. La ecuación de regresión X = -70870.43881+1.418630073e-005*PIB+1.11422759e-005*FBKF, en lo que se refiere a los signos de los coeficientes de las variables exógenas cumplen con lo especificado por la teoría del consumo.
b. Las pruebas de significación estadísticamente usando α=5% realizadas con t, para β_1, β_2 y F para R^2 tienen una probabilidad de casi cero, lo cual, corrobora que PIB y FBKF explican satisfactoriamente a X.
c. En ese sentido R^2= 0.960721 y la ajustada $\overline{R}^2$ = 0.960072 muestran que PIB, FBKF son suficientes para explicar a X; no se necesita otra variable para determinar el comportamiento presente y futuro de X.
d. Al igual que el ejercicio anterior se realizarán pruebas con el propósito de asegurarnos que la ecuación de regresión múltiple antes descrita sirva para

pronosticar y visualizar escenarios futuros, es decir, para hacer planeación sobre las exportaciones.

III. Pruebas de especificación del modelo.
1.- Omisión de variables explicativas, digamos INPC.

Para verificar si omitimos algún regresor explicativo en el modelo aplicamos la variable INPC en su “razón de verosimilitud” como el estadístico F, que permiten incorporar; uno o varios regresores explicativos y probar si su contribución al modelo es significativo estadísticamente. Así, planteamos con α=5%

Prueba de Hipótesis:

H_0: una o varias variable(s) explicativa(s) no es (son) significativa(s) estadísticamente. No debe(n) incorporarse al modelo: en este caso INPC = 0.

H_a: La variable independiente añadida es significativa, por tanto, debe incluirse en el modelo INPC ≠ 0.

Regla de decisión

a) Si la probabilidad es mayor que 0.05 se acepta H_0.

b) Si la probabilidad es menor o igual que 0.05 se rechaza la H_0.

Para evaluar la prueba de variables omitidas dentro de la pantalla de regresión en Eviews se selecciona dentro de la pantalla de regresión: view/coefficient tests/omitted variables- likelihood ratio/y en la caja de “omitted – redundant variable test, escribimos INPC/ok y obtenemos el siguiente cuadro:

Cuadro 8.3.25. Pantalla de Regresión sobre Variables Omitidas.

Omitted Variables: INPC

F-statistic	12.65393	Probability	0.000538
Log likelihood ratio	12.43125	Probability	0.000422

Test Equation:
Dependent Variable: X
Method: Least Squares
Date: 09/12/12 Time: 20:49
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	42616.96	8356.579	5.099809	0.0000
PIB	7.90E-06	2.02E-06	3.920395	0.0001
FBKF	1.29E-05	3.51E-06	3.684749	0.0003
INPC	257.1569	72.29175	3.557237	0.0005

R-squared	0.964468	Mean dependent var	33392.56
Adjusted R-squared	0.963579	S.D. dependent var	26035.35
S.E. of regression	4968.635	Akaike info criterion	19.89140
Sum squared resid	2.96E+09	Schwarz criterion	19.98238
Log likelihood	-1229.267	F-statistic	1085.736
Durbin-Watson stat	0.287230	Prob(F-statistic)	0.000000

En cuanto a la Prueba F que se encuentra en la regresión es de F= 1085.736, recordando en los ejemplos anteriores que esta prueba es la evaluación de la significación estadística en la regresión *X= f(PIB,FBKF)*. Con α= 5% observamos que la probabilidad de F (0.000538) y con Log Likelihoold Ratio (0.000422) < 0.05, entonces aceptamos la hipótesis alterna (H_a), por lo tanto, INPC es significativa estadísticamente y debemos incluirla en este modelo, replanteando el modelo con la variable omitida sería *X=f(PIB, FBKF,INPC)*

2. Variables explicativas redundantes, digamos PIB.

En Eviews la prueba “Redundant variables- likelihood ratio” contrasta la significación estadística de una o varias variables exógenas, con el fin de cerciorarse de que no sobran o de que son redundantes. Así con α=5% supóngase que

H_o: PIB es redundante

H_a: PIB no es redundante

Como en el ejemplo anterior partimos del cuadro de la ecuación de regresión en que *X=f(PIB, FBKF)*, sobre la pantalla de regresión seleccionamos view/coefficient tests/redundant variables-likelihood ratio/ escribimos “PIB” en “omitted redundant variables test” /ok y aparece el cuadro siguiente:

Cuadro 8.3.26. Pantalla de Regresión sobre Variables Redundantes.

Redundant Variables: PIB

F-statistic	195.4503	Probability	0.000000
Log likelihood ratio	119.2106	Probability	0.000000

Test Equation:
Dependent Variable: X
Method: Least Squares
Date: 09/13/12 Time: 18:05
Sample: 1980Q1 2010Q4
Included observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-38682.79	2332.620	-16.58341	0.0000
FBKF	5.95E-05	1.82E-06	32.64383	0.0000

R-squared	0.897273	Mean dependent var	33392.56
Adjusted R-squared	0.896431	S.D. dependent var	26035.35
S.E. of regression	8378.712	Akaike info criterion	20.92077
Sum squared resid	8.56E+09	Schwarz criterion	20.96626
Log likelihood	-1295.000	F-statistic	1065.620
Durbin-Watson stat	0.214504	Prob(F-statistic)	0.000000

Vemos que las probabilidades de F y de X^2 en la razón de verosimilitud son menores que 5%, por lo que aceptamos H_a y al no ser redundante el PIB, por lo tanto, la conservamos en el modelo como regresor explicativo de las Exportaciones (X).

IV. Pruebas de estabilidad estructural

a) Contraste de Chow

Se hace uso de esta prueba cuando se piensa que dentro de la serie existe un periodo que por diversas razones cambió la trayectoria de la variable endógena, en este caso con 126 observaciones (30 años representados en trimestres) suponga que la observación 11 colapsaron las exportaciones ya que fue en el primer trimestre de 1982, después de anunciada a la deuda de ese mismo año. Por el desequilibrio económico hay razones de peso para pensar que cambió el valor de las Exportaciones (X). Para verificarlo, con α=5% establecemos la hipótesis nula de estabilidad estructural.

H_o: Hay un solo modelo para todos los datos; probabilidad de F mayor que α=5%;

H_a: Cada subgrupo en que se dividen los datos tiene un comportamiento diferente: no hay un solo modelo hay dos; probabilidad de F menor que α=5%.

Luego si no hay diferencias estadísticas significativas entre el modelo restringido y el otro aceptamos Ho: de estabilidad estructural del modelo; en caso contrario, aceptamos Ha y decimos que hubo un cambio estructural en el periodo 11 en el año 1982 en el primer trimestre, en que todo el sistema económico se trastocó. Para aceptar o rechazar la prueba de hipótesis de la prueba de estabilidad Chow dentro de la pantalla de regresión señalamos view/stability tests/chow break point test/ damos click y en la ventana que aparece escribimos 1982.1/ok genera el siguiente cuadro:

Cuadro 8.3.27. Prueba de Estabilidad Chow.

Chow Breakpoint Test: 1982Q1			
F-statistic	1.097658	Probability	0.353031
Log likelihood ratio	3.413017	Probability	0.332220

Al ver que la probabilidad de la prueba-F es 0.353031>0.05, aceptamos H_0, es decir, hay un solo modelo para todos los datos.

b) Contraste de predicción de Chow.

Con α=5% planteamos:

H_0: Hay estabilidad estructural, en un modelo restringido a un solo comportamiento de todos los datos, cuyos residuos se comparan con los del período más largo de los dos en los que la serie ha sido dividida.

Partimos de la ecuación de regresión *X=f(PIB, FBKF),* se inicia el siguiente proceso: view/stability test/chow forecast tests/click y en “Chow test” escribimos el año en que pensamos a partir del cual pudo haber ocurrido un cambio estructural: 1982 / ok y aparece el cuadro:

Cuadro 8.3.28 Prueba de estabilidad de Chow: Pronóstico desde 1982Q1 a 2010Q4.

Chow Forecast Test: Forecast from 1982Q1 to 2010Q4

F-statistic	120.0151	Probability	0.000019
Log likelihood ratio	983.5851	Probability	0.000000

Test Equation:
Dependent Variable: X
Method: Least Squares
Date: 09/13/12 Time: 19:11
Sample: 1980Q1 1981Q4
Included observations: 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-5777.671	5215.838	-1.107504	0.3185
PIB	1.49E-06	2.41E-06	0.618681	0.5632
FBKF	5.42E-06	6.23E-06	0.870127	0.4240

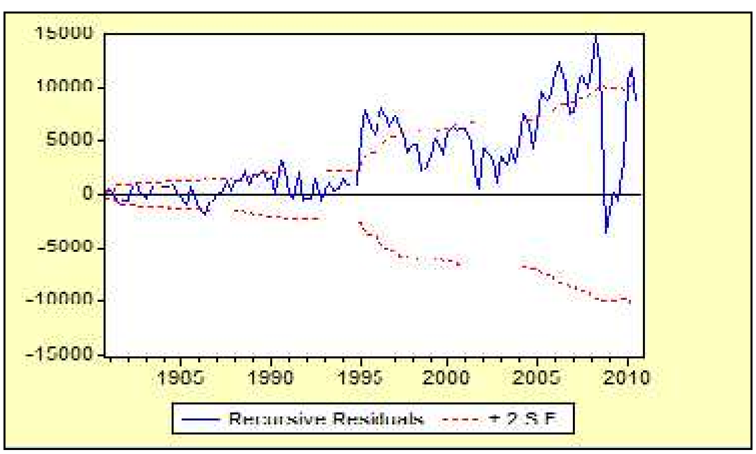
R-squared	0.811280	Mean dependent var	6996.523
Adjusted R-squared	0.735793	S.D. dependent var	943.4043
S.E. of regression	484.8203	Akaike info criterion	15.48584
Sum squared resid	1175739.	Schwarz criterion	15.51563
Log likelihood	-58.94337	F-statistic	10.74717
Durbin-Watson stat	1.740517	Prob(F-statistic)	0.015472

Rechazamos la H_0 ya que la probabilidad de los estadísticos es menor que 5% y concluimos diciendo que hay no estabilidad estructural dentro de esta muestra.

c) Estimación recursiva.

Como en los ejemplos anteriores, si no hay cambio estructural: Ho. “Las estimaciones de los parámetros se mantendrán constantes al ir aumentando las muestras secuencialmente y los residuos no se desviarán amplia o significativamente de cero” Así, de View/stability tests/Recursive Estimate (OLS only)/ok y en la caja de diálogo por default aceptamos C(1) C(2) / ok y aparece esta gráfica:

Gráfica 8.3.6. Estimación recursiva.



Podemos apreciar dentro de la gráfica que no hay estabilidad en el modelo en el período. Los residuos recursivos se salen de las bandas construidas con dos errores estándar alrededor de Y ± 2S.E.

d) Errores de especificación en la forma funcional/ RESET

Con α=5% la probabilidad de rechazar una hipótesis cierta, establecemos;

H_o: Hay linealidad en el modelo

H_a: No hay linealidad en el modelo

Dar View/Stability tests/Ramsey Reset Test/ clic y aparece el cuadro de diálogo "RESET Specification" y escribimos el *número de potencias* de la variable endógena ajustada a incluir empezando por el cuadrado; así si indicamos se añadirá el cuadrado de dicha variable, si ponemos 2 se incluirá el cuadrado y el cubo, etc. Si dejamos la celda en blanco el programa entiende que se añadirá la variable al cuadrado.

El resultado son las variables de F y χ^2 de razón de verosimilitud junto con la ecuación estimada. Luego de View/ Stability tests/Ramsey RESET y escribimos en la celda en blanco 1 /ok y aparece el cuadro.

Cuadro 8.3.29. Prueba Ramsey RESET.

Ramsey RESET Test:

F-statistic	253.1081	Probability	0.000000
Log likelihood ratio	140.6627	Probability	0.000000

Test Equation:
Dependent Variable: X
Method: Least Squares
Date: 09/13/12 Time: 20:12
Sample: 1980Q1 2010Q4
Included observations: 121

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	21972.05	3441.319	6.383847	0.0000
PIB	8.32E-06	6.86E-07	12.13664	0.0000
FBKF	-1.14E-05	2.51E-06	-4.547593	0.0000
FITTED^2	9.07E-06	5.70E-07	15.90937	0.0000

R-squared	0.987367	Mean dependent var	33392.56
Adjusted R-squared	0.987051	S.D. dependent var	26035.35
S.E. of regression	2962.646	Akaike info criterion	18.85728
Sum squared resid	1.05E+09	Schwarz criterion	18.94326
Log likelihood	-1165.151	F-statistic	3126.292
Durbin-Watson stat	1.337517	Prob(F-statistic)	0.000000

Al tener F y χ^2 de razón de verosimilitud probabilidades menores que 5% rechazamos H_o y concluimos que el modelo no es lineal.

V. Normalidad entre las perturbaciones

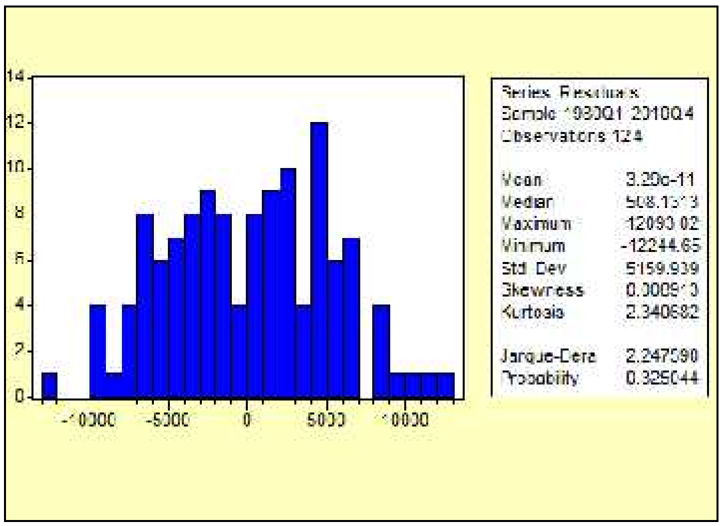
Para ello es fundamental plantear con α=5% y verificar:

H_0: Hay normalidad en las perturbaciones: JB=0

H_a: No hay normalidad en las perturbaciones: JB $\neq$ 0

Como no son observables las perturbaciones se estudian con los residuos. Si verificamos Ho ello indica que la distribución empírica de los residuos debe ser similar a la de la distribución normal. De View/Residual tests/Histogram normality test/ click y aparecen el cuadro y la gráfica siguientes:

Gráfica 8.3.7. Histograma.



Con la probabilidad de JB=0.325044 >0.05 y JB se aproxima a cero aceptamos Ho; además de que la kurtosis se acerca a 3, a pesar de que la asimetría no sea cero.

VI. Predicción.

En esta parte es importante señalar que los datos que se estimaron para los trimestres de 2011.1 a 2011.4, en la variables de Producto Interno Bruto (PIB) y Formación Bruta de Capital Fijo (FBKF), se utilizó el método de Suavizamiento simple de Holt, el cual se verá más adelante.

6. Partimos del archivo que hemos estado trabajando desde el inicio de este ejemplo, seleccionamos la regresión X=f(PIB, FBKF) .
7. Expandimos en 1 año el rango del archivo, es decir, en cuatro trimestre que van de 2011.1 a 2011.4, seleccionamos de la barra de herramientas que se encuentra arriba de nuestra barra de comandos Proc/ Structure-Resize Current…/clic/ End date: 2011Q4/ ok
8. Damos en la ventana de la ecuación estimate: procs/make regresión group, aparecen los datos de los trimestre 2011.1 a 2011.4 en X, PIB, FBKF en blanco.

Editamos seleccionando en la barra de herramientas del grupo de variables “edit+/-“, escribir sus valores proyectados previamente obtenidos con el método de Suavizamiento de Holt, los cuales pueden apreciarse en el siguiente cuadro. Tenemos las siguientes tablas, solo se muestra los últimos datos, es decir, de 2009 1 a 2011.4.

Tabla 8.3.7 Datos

Obs	X	PIB	FBKF
2009Q1	61396.88	8071420150	1817554845
2009Q2	64265.02	8177723778	1794339075
2009Q3	68526.08	8512165089	1833709460
2009Q4	77222.46	8833689500	1888408924
2010Q1	78101.52	8434182477	1771876167
2010Q2	85464.74	8807669828	1827505452
2010Q3	85776.04	8961898776	1901021623
2010Q4	91441.89	9239060458	2005574203
2011Q1			
2011Q2			
2011Q3			
2011Q4			

Para llenar las celdas de PIB, FBKF como se mencionó en los dos ejemplos anteriores, existen dos procedimientos:

3. Escribimos los valores proyectados de PIB, FBKF (que son conocidos con antelación) como en el siguiente cuadro (llamado grupo 2) mismo que lo guardamos como grupo2.
4. En la línea de comando escribir Data PIB FBKF/ enter y aparece el cuadro del grupo 4 y llenamos los años con los datos correspondientes.

Grupo 2

Tabla 8.3.8 Datos

Obs	X	PIB	FBKF
2009Q1	61396.88	8071420150	1817554845
2009Q2	64265.02	8177723778	1794339075
2009Q3	68526.08	8512165089	1833709460
2009Q4	77222.46	8833689500	1888408924
2010Q1	78101.52	8434182477	1771876167
2010Q2	85464.74	8807669828	1827505452
2010Q3	85776.04	8961898776	1901021623
2010Q4	91441.89	9239060458	2005574203
2011Q1		2013874556	9134012795
2011Q2		2027843563	9179787610
2011Q3		2041812571	9225562426

2011Q4		2055781578	9271337241

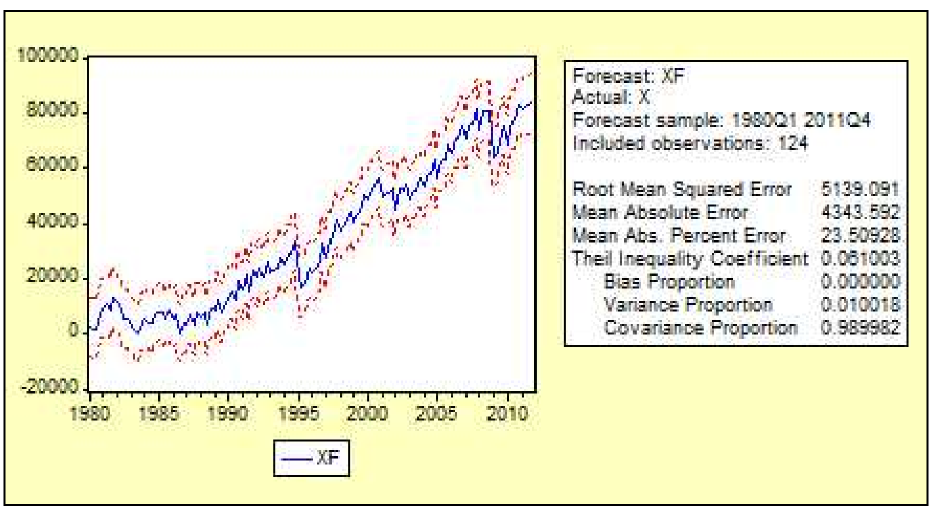
Grupo 4

Tabla 8.3.9 Datos

Obs	PIB	FBKF
2009Q1	8071420150	1817554845
2009Q2	8177723778	1794339075
2009Q3	8512165089	1833709460
2009Q4	8833689500	1888408924
2010Q1	8434182477	1771876167
2010Q2	8807669828	1827505452
2010Q3	8961898776	1901021623
2010Q4	9239060458	2005574203
2011Q1	2013874556	9134012795
2011Q2	2027843563	9179787610
2011Q3	2041812571	9225562426
2011Q4	2055781578	9271337241

Evaluación: Regreso a la ecuación original la que hemos estado analizando, nuevamente dentro de la barra de herramientas esta la opción de Forecast /aparece una pantalla que muestra una nueva variable XF (variable endógena pronosticada), dentro del cuadro "Forecast sample", que va en el periodo de 1980q1 a 2011q4/ ok y aparece el siguiente cuadro:

Grafico 8.3.8 Predicción



Cuyos resultados estadísticos del lado derecho sirven para evaluar la calidad de la predicción, mismos que se interpretarán párrafos abajo.

Ahora en la pantalla de workfile: ecuación para predecir, seleccione (sombreo) X y XF para conocer el valor pronosticado de XF/ open group/clic y aparece el siguiente cuadro que llamo

Este cuadro también se obtendría si sobre la barra de comandos escribimos “Show X XF”/ enter y aparece igual al caso anterior.

Grupo 03.

Tabla 8.3.10 Datos

obs	X	XF
2009Q1	61396.88	63884.8523
2009Q2	64265.02	65134.231
2009Q3	68526.08	70317.3917
2009Q4	77222.46	75488.1102
2010Q1	78101.52	68522.1433
2010Q2	85464.74	74440.384
2010Q3	85776.04	77447.4597
2010Q4	91441.89	82544.3124
2011Q1		81146.5595
2011Q2		81951.5813
2011Q3		82756.6031
2011Q4		83561.6249

Evaluación de la Predicción: Puesto que ya conocemos XF, ahora podemos evaluar estadísticamente dicha predicción analizando en la figura anterior la gráfica y las estadísticas del cuadro. En el caso de la gráfica observamos dos caídas de importancia en nuestra serie debido a los periodos de recesión en la economía mexicana los cuales fueron el error de diciembre en 1994-1995 y la crisis de sistema inmobiliario que se originó en Estados Unidos, afectando a los mercados financieros y reales de todo el mundo.

En este ejercicio analizando los resultados estadísticos, probablemente no sea muy bueno el pronóstico, dado que la covarianza es cercana a uno. Por consiguiente, para predecir, no resultó adecuada la ecuación de regresión X=f(PIB, FBKF), probablemente el método de Suavizamiento de Holt para pronosticar los datos futuros no haya sido el correcto o la forma funcional sea incorrecta.

CAPITULO IX.- NÚMEROS ÍNDICES

Su relación con las serie de tiempo.

Ambos miden fluctuaciones de la variable económica (precios, cantidades, valor, consumo, inflación, inversión) en el tiempo. Ambos permiten pasar de un análisis estático a uno dinámico al meter la variable tiempo de manera que ésta actúa como variable independiente o explicativa de la variable económica, que así se convierte en variable dependiente.

Diferencias, semejanzas y uso: Los *índices tradicionalmente conocidos* son los de los precios, Ip, los de cantidades, Iq, y los de valor, Iv, cuya característica común es que miden las fluctuaciones de un momento en el tiempo llamado "base" a otro denominado "de comparación", es decir, miden las variaciones entre dos puntos. Al ser de índole *temporal* las fluctuaciones, éstas tienen *periodicidad*, es decir, pueden ser diarias, semanales, mensuales, anuales, etc. Así, con esta característica, como se verá en la segunda parte de este capítulo, algunos de estos índices suelen *usarse* en la identificación y eliminación de "irregularidades" observadas en las series de Tiempo.

Las series de tiempo son registros de las oscilaciones de los valores de una variable, Y, en el tiempo, o sea que éste denotado con t, actúa como la variable independiente que explica los cambios de Y_t en el tiempo, generalmente para muchos más momentos o puntos de Y_t en el tiempo. Así, al graficar sus fluctuaciones, éstas pueden mostrar *efectos* circunstanciales, tendenciales o estacionales en ciertos momentos en el tiempo, mismos que deben de eliminarse con objeto de mostrar la dirección "real" de la variable económica.

Por su característica intrínseca de temporalidad de sus datos, para su análisis se debe de partir de la consideración de que los datos del pasado influyen en el dato actual y por consiguiente en sus proyecciones futuras, independientemente de que dichas proyecciones se hagan con métodos de econometría tradicional (análisis bivariado) o de la moderna o auto regresiva (análisis univariado).

Alcance del análisis de las series de tiempo: su estudio generalmente sirve para hacer análisis estructural, pero es indudable que también permite evaluar políticas económicas (públicas o privadas) y para predecir, dado que al limpiar o librar Yt de cualesquiera de alguno de los efectos antes mencionados, *ahora* se está en condiciones de poder *predecir o proyectar* su comportamiento futuro con mayor certeza (porque sus oscilaciones en el tiempo han sido suavizadas con técnicas como las de los índices, diferencias, medias móviles, logaritmos, etc. que veremos enseguida) que *antes* cuando estaba afectada por aspectos circunstanciales, tendenciales, cíclicos o estacionales. Así, al familiarizarse el investigador con los métodos de análisis, está en condiciones de comparar los valores de Y_t antes y después de aplicar la nueva política pública (ergo, fiscal) o política privada (mercadológica).

Derivado de lo anterior es por ello que en la actualidad se ha masificado el estudio de los métodos de las series de tiempo, tanto con los de la llamada “Econometría tradicional” como con los de “Series de tiempo o moderna”. Al respecto, los aspectos básicos de los primeros se vieron en los primeros 10 capítulos anteriores, mismos que ahora sirven de referencia fundamental para en este momento iniciar el estudio de los segundos.

IX.1.-Números índices.

Un índice, I, mide la variación de un fenómeno generalmente a través del tiempo. En Economía se usan mucho y en particular para medir las variaciones de los precios, de las cantidades y del valor de los bienes y servicios que existen en el mercado de uno a otro periodo. Aparte de medir las variaciones de los precios, Ip, de las cantidades, Iq, y del valor,Iv, un índice es muy útil porque expresa el cambio del *grupo heterogéneo* de bienes y servicios que integran la *muestra* (automóviles, frijoles, camisas, televisores, corbatas, cepillos de dientes, etc.) que se usa para calcularlos, es decir, es muy práctico convertir sus datos en índices para cuantificar sus variaciones en el tiempo.

Origen

En opinión de Mason (2001: 620), al italiano G.R. Carli se le atribuye la paternidad de estos indicadores, ya que los elaboró e integró a un informe que hizo en el año de 1764 sobre las oscilaciones de los precios de Europa de 1500 a 1750.

¿Por qué se acostumbra convertir los datos originales en índices? Porque sólo así se pueden comparar los precios y cantidades de grupos de bienes y servicios de diversa índole (zapatos, mantequilla, medicinas, automóviles, etc).

El índice es una medición hecha sobre variaciones en el tiempo de los precios, cantidades o valor de uno o varios bienes y servicios existentes en el mercado. Por convención se toma una base para medir esa variación tomando como referencia 100%; de tal manera que cuando el índice por ejemplo es 83%, ello significa que hubo una disminución del 17%, de igual manera cuando es digamos, 325%, ello indica que hubo un aumento de 225%.

Los índices tienen una gran aplicación, en la actualidad constituyen la columna vertebral para la toma de decisiones en el combate a la inflación, para medir la productividad de los factores de la producción y para medir la rentabilidad de las inversiones, entre otras aplicaciones.

Los índices son de diferente naturaleza; su cálculo se basa en el muestreo estadístico debido a la amplia gama de bienes y servicios existentes en el universo económico, por lo que se opta para calcularlos utilizando un reducido número de ellos, es decir, una muestra proveniente del vasto universo compuesto por los bienes y servicios existentes en el mercado en un momento dado .

IX.1.1.-Tipos de índices

Los hay relativos/simples y compuestos o ponderados. Nosotros vamos a calcular unos y otros para los precios, las cantidades y de valor.

IX.1.1.1.-Números índices simples o relativos.

Los números índice relativos son los porcentajes que expresan variaciones de precio o cantidad de un producto X (en comparación con su precio o cantidad de un año base). También estas variaciones pueden ser de varios productos o servicios.

Para calcular números índices de precios se requiere: seleccionar los artículos, selección del período base (se sugiera que debe ser después de un censo), de los precios de los artículos y selección de la fórmula.

Puesto que una variación se mide en el tiempo, llamaremos P_o y Q_o a los precios y cantidades, del año, (día o mes) base o de referencia, y P_1, Q_1 a los precios y cantidades del año (día o mes) de comparación. Así, una variación en términos relativos será:

$$I_p = \frac{P_1}{P_0} * 100; \quad I_q = \frac{Q_1}{Q_o} * 100$$

I_p é I_q indican el índice de precios y cantidades, respectivamente.

Un índice relativo se puede calcular para una mercancía o servicio, como el caso anterior o para varios, como sucede en la realidad.

Su fórmula es:

$$I_p = \frac{\sum P_1}{\sum P_0} * 100; \quad I_q = \frac{\sum q_1}{\sum q_0} * 100$$

Igualmente, los índices simples ó relativos como promedios que pretenden ser representativos, de las variaciones de los fenómenos suelen calcularse con las siguientes fórmulas, según la naturaleza y características de los fenómenos.

Media aritmética:

$$I_p = \frac{\sum \frac{P_1}{P_0}}{n} * 100; \quad I_q = \frac{\sum \frac{q_1}{q_0}}{n} * 100$$

Media geométrica Log:

$$I_p = \frac{\sum \log \frac{P_1}{P_0}}{n} + \log 100 - \log n$$

Media geométrica Log:

$$I_q = \frac{\sum \log \frac{q_1}{q_0}}{n} + \log 100 - \log n$$

Media armónica:

$$\boldsymbol{I_p} = \frac{\boldsymbol{n}}{\sum \frac{\boldsymbol{P_i}}{\boldsymbol{P_o}}} * 100 \qquad \boldsymbol{I_q} = \frac{\boldsymbol{n}}{\sum \frac{\boldsymbol{q_1}}{\boldsymbol{q_o}}} * 100$$

Estadísticamente estas fórmulas expresan promedios, en este caso de las variaciones. Por consiguiente las limitaciones que tiene la media aritmética de que es afectada por los valores extremos incidenen en estos índices (relativos) que por consiguiente no miden objetivamente las variaciones, por lo que su uso es limitado (cuando los datos son homogéneos). *Para superar este limitante se usan factores de ponderación en la forma que se demuestra a continuación.*

IX.1.1.2.- Números índices compuestos o ponderados.

En el índice de precios el factor de ponderación es la *cantidad* y en el índice de cantidades el factor de ponderación es el *precio*.

Luego:

$$I_p = \frac{\sum P_1 Q}{\sum P_0 Q}; \quad Iq = \frac{\sum q_1 P}{\sum q_0 P}$$

Al respecto, el factor de ponderación puede ser el del año base o el del año de comparación. Cuando es el año base, la fórmula es:

$$I_p = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} * 100; \quad Iq = \frac{\sum Q_1 P_0}{\sum Q_0 P_0} * 100$$

que elaboró Laspeyres. Cuando es el año de comparación se usan las fórmulas elaboradas por Paasche:

$$I_p = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} * 100; \quad Iq = \frac{\sum q_1 P_1}{\sum q_0 P_1} * 100$$

En este sentido el experto Fisher formula una ponderación de las dos anteriores y la llamo: **"Fórmula ideal de Fisher"**, la cual viene dada por:

$$I_p = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} * \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} * 100$$

$$I_q = \sqrt{\frac{\sum q_1 P_0}{\sum q_0 P_0} * \frac{\sum q_1 P_1}{\sum q_0 P_1}} * 100$$

Derivado de los desarrollos anteriores podemos decir que el índice del valor se calcula con la siguiente formula:

$$\boldsymbol{Iv} = \frac{\sum \boldsymbol{P}_1 \boldsymbol{q}_1}{\sum \boldsymbol{P}_0 \boldsymbol{q}_0} * 100$$

IX.1.1.3.- Pruebas matemáticas para escoger el índice más apropiado.

Puesto que hay diferentes métodos para calcular índices (Marshall, Keynes, Ellsworth, etc.) el experto Irving Fisher ideó una serie de criterios matemáticos para que con base en ellos el invstigador pudiera seleccionar el más adecuado para medir las variaciones en el tiempo: de precios, cantidades o de valor. Así, a continuación se muestran algunos criterios matemáticos utilizados para seleccionar el índice más apropiado.

- Reversión cronológicas; y
- Reversión de factores.

De tal suerte que el índice que pasa "esas" pruebas matemáticas" es el que debe usarse en opinión de Fisher. Como se verá más adelante con un ejemplo numérico, sólo el índice de Fisher pasa estas pruebas, por eso lo llamó "ideal".

IX.1.1. 4.- Cambio de base.

Es algo que fácilmente y de manera rutinaria el investigador suele hacer, en particular cuando la serie es ya demasiado larga.

Ello significa que el cambio de base se hace por comodidad, ergo; por ello expresa las variaciones en función de un año reciente, pero que de ninguna manera mejora la serie o valores del fenómeno bajo estudio.

IX.1.1.5.- Deflactación.

El índice de deflactación sirve para expresar en términos reales, es decir, a precios constantes de un año base seleccionando previamente, las variaciones de un fenómeno económico; ergo; ingreso, salario, ventas, es decir, por medio del proceso de deflactación se elimina el efecto de los precios en el análisis de un fenómeno económico (salario, ingreso, ventas) para que este quede expresado en forma real y la medición de sus variaciones sea objetiva y no distorsionada por los precios corrientes de los bienes y servicios.

Para deflactar los datos de un fenómeno económico, lo que se hace primero es seleccionar el deflactor o índice correspondiente a la naturaleza de ese fenómeno. Al respecto es conveniente señalar que en México se calculan diversos índices de precios de los cuales destacan: el índice de precios al consumidor, al productor, la vivienda, PIB, etc.

Una vez seleccionado el deflactor, para convertir valores nominales (o precios de mercado) en valores reales (a precios constantes de un año base seleccionado previamente) se divide cada uno de los datos, entre el valor del deflactor en ese año.

Así se hace para todos los datos del fenómeno bajo estudio durante un período de tiempo determinado. El cociente resultante es el valor real, en cada año, del fenómeno de interés.

Por analogía, conservando el espíritu de eliminar el efecto de los precios, también se puede inflactar a precios reales los valores del fenómenos de interés.

El profesor Alberto Reyes de la Rosa homogenizó la información al deflactar de 1968 al año 2002, como se expone a continuación.

Tabla 9.1.1.5.1 Deflactación.

1	2	3	4	5	6	7
Base 1968=100	INPC	**Base 1978=100**	INPC	**Base 1994=100**	INPC	**Inflación Base 2002=100**
1968	**100**	30.2	30.2	0.08	0.08	
1969	103.5	31.3	31.3	0.08	0.08	3.5
1970	108.7	32.9	32.30	0.09	0.09	3.2
1971	114.6	34.6	34.00	0.09	0.09	5.3
1972	120.3	36.4	35.70	0.10	0.10	5.0
1973	134.8	40.7	40.00	0.11	0.11	12.0
1974	166.8	50.4	49.50	0.13	0.13	23.8
1975	191.8	58.0	57.00	0.15	0.15	15.2
1976	222.1	67.1	66.00	0.18	0.18	15.8
1977	286.7	86.7	85.10	0.23	0.23	28.9
1978	330.8	**100.0**	**100.00**	0.27	0.27	17.5
1979	117.8	35.6	118.20	0.32	0.32	18.2
1980	149.0	45.0	149.30	0.40	0.40	26.3
1981	191.9	58.0	191.10	0.51	0.51	28.0
1982	302.4	91.4	303.60	0.81	0.81	58.9
1983			612.90	1.64	1.64	101.9
1984			1,014.10	2.71	2.71	65.5
1985			1,599.70	4.28	4.27	57.7
1986			2,979.20	7.97	7.95	86.2
1987			6,906.60	18.47	18.43	131.8
1988			14,791.20	39.55	39.47	114.2
1989			17,705.60	47.35	47.25	19.7
1990			22,481.50	60.12	60.00	27.0
1991			27,576.30	73.75	73.59	22.7
1992			31,852.80	85.18	85.01	15.5
1993			34,959.00	93.49	93.29	9.8
1994			37,394.10	**100.00**	**99.79**	7.0
1995			50,478.30	134.99	134.71	35.0
1996			67,836.64	181.41	181.04	34.4
1997			81,828.39	218.83	218.37	20.6
1998			94,890.15	253.76	253.23	16.0
1999			110,595.67	295.76	295.15	16.6
2000			121,092.62	323.83	323.16	9.5
2001			128,187.35	342.80	342.09	5.9
2002				**100.21**	**100.21**	5.7

Se parte inicialmente de los datos que se obtenien en de la fuente de información que es la columna número 2, base 1968 = 100, para pasar de la base 1968 a 1978=100 es necesario realizar una simple operación aritmética que es la división de $\frac{100}{330.8}*100 = 30.2$, el dato de 330.8 se usa por ser el año al que se va a "arrastrar la información", para el siguiente año la operación es $\frac{103.5}{330.8}*100 = 31.3$ y asi sucesivamente hasta donde se desea hacer el cambio de base.

En la columna 4 es solamente el INPC con base 1978, para cambiar la base a 1994 los resultados aparecen en la columna 5, los calculos son los siguientes para el año 1968 $\frac{30.2}{37,394.10}*100 = 0.0807$, para el año 1975

IX.1.1.6.-CONSTRUCCIÓN DE ÍNDICES

1.- ÍNDICES RELATIVOS PARA UN SOLO ARTÍCULO

Un vendedor de refrigeradores tiene las siguientes ventas:

Tabla 9.1.1.6.1 Construcción de Índices

Año	Precio	No. de	Ingresos en
	Prom. c/u	Unidades	Miles
1	2	3	3*2
1996	3,000	60	180
1997	3,300	63	207.9
1998	3,900	60	234
1999	4,500	66	297
2000	4,500	72	324
2001	4,800	75	360
2002	4,950	66	326.7

Considerando 1996 = 100, es decir, año base, los índices se calculan así:

Tabla 9.1.1.6.2 Construcción de Índices

Año	Precio		Cantidad		Ingresos	
	$	Índice	Unidades	Índice	$·	Índice
1996	3,000	100	60	100	180	100
1997	3,300	110	63	105	207.9	116
1998	3,900	130	60	100	234	130

1999	4,500	150	66	110	297	165
2000	4,500	150	72	120	324	180
2001	4,800	160	75	125	360	200
2002	4,950	165	66	110	326	181.11

Si ahora cambiamos de base digamos al año 2000 = 100, haciendo los cálculos con dos procedimientos para los precios tendremos:

Tabla 9.1.1.6.3 Construcción de Índices

Año	Índice Anterior	Índice con Números Originales	Nuevo obtenido con Índices
1996	100	3,000 ÷ 4500 = 67	100 ÷ 150 = 67
1997	110	3,300 ÷ 4500 = 73	110 ÷ 150 = 73
1998	130	3,900 ÷ 4500 = 87	130 ÷ 150 = 87
1999	150	4,500 ÷ 4500 = 100	150 ÷ 150 = 100
2000	150	4,500 ÷ 4500 = 100	150 ÷ 150 = 100
2001	160	4,800 ÷ 4500 = 107	160 ÷ 150 = 107
2002	165	4,950 ÷ 4500 = 110	165 ÷ 150 = 110

Lo mismo puede hacerse para las CANTIDADES y los INGRESOS.

IX.1.1.7.- Aplicaciones para deflactar e inflactar.

La deflactación se hace lo mismo para una serie cronológica como para el análisis comparativo en dos años de un fenómeno en términos reales.

Así por ejemplo, si deseamos conocer el ingreso real de una persona digamos de 2010 a 2011, tomando en cuenta que el primer año su ingreso nominal fue de $10 millones y en el segundo fue de $12.6 millones. El procedimiento es el siguiente.

Con 2010 = 100%

Año	Salario Nominal	Ip	Ingreso Real en miles
2010	$10 millones	100	$\frac{\text{Ingreso Nominal}}{\text{Ip}} = \frac{10}{1} = 10$
2011	$12.6 millones	110	$\frac{\text{Ingreso Nominal}}{\text{Ip}} = \frac{12.6}{1.10} = 11.45$

En ocasiones es necesario INFLACTAR los valores de un fenómeno económico, como las ventas anuales de una empresa.

Deflactar = Quitarle el efecto de los precios.

Inflactar = Aumentar el efecto de los precios.

Coeficiente = Valor real.

Por ejemplo, en 2011 se deseaba *inflactar* las ventas hechas por las empresas durante 2008, 2009, 2010 y 2011.

Para ello se cuenta con el índice de precios al consumidor para esos años el cual nos permite mediante el cambio de base, hacer la inflactación correspondiente: 2011 = 100

9.1.1.7.1 Aplicaciones para deflactar e inflactar.

Año	Índice	Nuevo Índice	
		Para dividir	Para multiplicar
2011	153.63	153.63 ÷ 153.63 =100	153.63 ÷ 153.63 = 100
2010	118.18	118.18 ÷ 153.63 =0.77	153.63 ÷ 118.18 = 1.3
2009	99.95	99.95 ÷ 153.63 =0.65	153.63 ÷ 99.95 = 1.54
2008	85.1	85.1 ÷ 153.63 =0.55	153.63 ÷ 85.1 = 1.82

Ejemplo $100 millones de ventas de 2008, 2009 y 2010 equivalen a precios de 2011 a:

9.1.1.7.2 Aplicaciones para deflactar e inflactar.

Año	Ventas				(Millones de pesos de cada empresa)				
2008	$100	÷	0.55	=	$ 182 mil	=	$100	*	1.82
2009	$100	÷	0.65	=	$ 154 mil	=	$100	*	1.54
2010	$100	÷	0.77	=	$ 130 mil	=	$100	*	1.3

Ahora bien para deflactar.- Si fijamos 2008 = 100 año base, es decir, llevamos el valor de las ventas a precios de 1999, en este caso se hace lo contrario, es decir, hacemos un cambio de base al revés.

9.1.1.7.3 Aplicaciones para deflactar e inflactar.

Año	Índice Anterior		Nuevo Índice									
	Dividir	Multiplicar	Para dividir					Para Multiplicar				
1999	0.55	1.82	0.55	÷	KRR=	=	1	1.82	÷	1.82	=	1
2000	0.65	1.54	0.65	÷	KRR=	=	1.18	1.54	÷	1.82	=	0.85
2001	0.77	1.3	0.77	÷	KRR=	=	1.4	1.3	÷	1.82	=	0.71
2002	1.00	1	1	÷	KRR=	=	1.81	1	÷	1.82	=	0.55

Así $100 millones de 1999, 2000, 2001, y 2002 equivalen a precios de 1999 a:

9.1.1.7.4 Aplicaciones para deflactar e inflactar.

Año	Millones de $ en ventas de cada empresa								
1999	100	÷	1	=	100	=	100	*	1.00
2000	100	÷	1.18	=	85	=	100	*	0.85
2001	100	÷	1.4	=	71	=	100	*	0.71
2002	100	÷	1.81	=	55	=	100	*	0.55

De los cálculos anteriores se puede deducir um indicador muy útil y por consiguiente muy usado en economia, el cual es el siguiente:

PODER ADQUISITIVO = 1÷ I_P

IX.1.1.8.- Caso real: Cálculo de la inflación mensual acumulada en México.

A continuación se muestran los cálculos que hace el Banco de México para determinar el índice inflacionario mensualmente. Aun cuando el ejemplo se refiere al año de 1990, la metodología está vigente.

Cálculo de la tasa de inflación acumulada a partir de las tasas mensuales de inflación. Para ello se toma como referencia el I.N.P.C. Índice Nacional de Precios al Consumidor con 1978 = 100 así para 1990:

Tabla 9.1.1.8.1 Cálculo de la inflación mensual acumulada en México.

Mes		Índice Nacional de Precios al Consumidor	I Variación Mensual Del INPC	II Base inicial para aplicar la inflación del mes (100+col. IV del renglón anterior)	III Importe de la Inflación del mes %	IV Inflación Acumulada %
A	Enero	20,260.7	4.8	100	4.8	4.8
B	Febrero	20,719.5	2.3	104.8	2.4104	7.2104
C	Marzo	21,084.8	1.8	107.2104	1.92978	9.14018
D	Abril	21,405.7	1.5	109.141187	1.63712	10.7773
E	Mayo	21,779.2	1.7	110.778305	1.88323	12.66053
F	Junio	22,258.9	2.2	112.661536	2.47855	15.13908
G	Julio	22,664.8	1.8	115.14009	2.072522	17.211602
H	Agosto	23,051.0	1.7	117.212612	1.992614	19.204216
H	Septiembre	23,379.6	1.4	119.205226	1.668873	20.873089
J	Octubre	23,715.7	1.4	120.874099	1.692237	22.565326

	20.6	22.565326

Para obtener la tasa mensual acumulada, no se debe sumar las tasas de inflación de cada mes, se debe multiplicar y después sumar; para así acumular correctamente las tasas de inflación de cada mes.

Así al empezar el mes de enero de 1990, se parte de la base 100 (columna I renglón A). La tasa de inflación del mes de enero fue de 4.8% luego la tasa de inflación acumulada al final del mes fue del 4.8 (columna IV renglón A)

La tasa de inflación del mes de febrero fue de 2.3 %. Sin embargo la tasa de inflación acumulada durante estos dos meses de 1990 no fue la simple suma de 4.8 + 2.3 = 7.1. El cálculo de la inflación acumulada al 29 de febrero es :

104.8 x 0.023 = 2.4104 + 4.8 = 7.21 % (columna IV renglón B).

GENERALIZANDO PARA:

Tabla 9.1.1.8.2 Cálculo de la inflación mensual acumulada en México.

Marzo	107.2104	*	0.018	=	1.92978	+ 7.2101	=	9.14018
Abril	109.141187	*	0.015	=	1.63712	+ 9.14016	=	10.7773
Mayo	110.778305	*	0.017	=	1.88323	+ 10.7783	=	12.66053
Junio	112.661536	*	0.022	=	2.47855	+ 12.66165	=	15.13908
Julio	115.14009	*	0.018	=	2.072522	+ 15.14009	=	17.211602
Agosto	117.212612	*	0.017	=	1.992614	+ 17.212612	=	19.204216
Septiembre	119.205226	*	0.014	=	1.668873	+ 19.205226	=	20.873089
Octubre	120.874099	*	0.014	=	1.692237	+ 20.874099	=	22.565326

Puede observarse que al finalizar el mes de octubre de 1990, la tasa de , inflación fue del 22.565326 (columna III y columna IV)y no del 20.6 (columna I) como lo indicaría simplemente la suma de las tasa de inflación mensual.

IX.1.1.9.- Ejemplos adicionales.

IX.1.1.9.1.-Para números relativos simples, reiteración:

Media de relativos Mg de relativos

$$I_p = \frac{\sum \frac{P_1}{P_0} * 100}{n} \qquad\qquad I_p = \frac{\sum \frac{P_1}{P_0}}{n}$$

log Ip= $\log\left[\frac{\sum \frac{P_1}{P_0} * 100}{n}\right] = \log I_p = \sum \frac{P_1}{p_0} + \log 100 - \log n$

log Ip= $\log\left[\frac{\sum \frac{P_1}{P_0}}{n}\right] + \log 100$

Ma de los relativos $I_p = \frac{n}{\sum \frac{P_0}{P_1} * 100}$

IX.1.1.9.2.-Para números índices compuestos o ponderados:

Si designamos con Q_o y Q_1 las cantidades de cada articulo del año base y del año de estudio y designamos ***IP*** el índice de precios tenemos que: $\sum P_o Q_o$ representa el valor total o costo total de bienes y servicios o artículos en el año base. La cantidad $\Sigma P_1 Q_0$ representa el valor total del mismo conjunto de artículos en año de comparación; entonces el índice de Laspeyres nos sirve para medir el costo total en cualquier año de comparación de un conjunto fijo de artículos comprados en el año base.

En el cálculo del índice Paasche $\Sigma P_0 Q_1$ es el valor total o costo total de artículos comprados en el año de comparación considerando los precios del año base y; $\Sigma P_1 Q_1$ es el valor total de artículos comprados en el año de comparación a los precios de ese mismo año. Entonces el índice de Paasche nos sirve para medir el costo total de los artículos en el año de comparación relativos al costo que podrán haber sido comprados si se hubiera adquirido en el año base.

Laspeyres.

Método del año base $I_p = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} * 100$

Paasche

Método del año de comparación $I_p = I_p = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} * 100$

Marshall:

$$I_p = \frac{\sum P_1 (Q_0 + Q_1)}{\sum P_0 (Q_0 + Q_1)} * 100$$

Fisher:

$$I_p = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} * \frac{\sum Q_1 P_1}{\sum Q_0 P_1}} * 100$$

$$I_q = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} * \frac{\sum Q_1 P_1}{\sum Q_0 P_1}} * 100$$

donde Iq = índice de cantidad

Iv = índice de valor

Valor:

$$Iv = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} * 100$$

Tabla 9.1.1.9.2.1 Para números índices compuestos o ponderados

		2001		2002					
Artículo	Unidad	Po	Qo	P1	Q1	P1Qo	PoQo	P1Q1	PoQ1
Maíz	Kgs.	2	3	3	1	9	6	3	2
Arroz	Kgs.	4	3	6	2	18	12	12	8
Papa	Kgs.	6	4	9	3	36	24	27	18
Trigo	Kgs.	8	5	12	4	60	40	48	32
Sal	Kgs.	10	6	15	5	90	60	75	50
		30	21	45	15	213	142	165	110

Laspeyres.

$$I_p = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} * 100 = \frac{213}{142} * 100 = 1.5 * 100 = 150\%$$

Paasche

$$\boldsymbol{I_p} = \frac{\sum \boldsymbol{P_1 Q_1}}{\sum \boldsymbol{P_0 Q_1}} * 100 = \frac{165}{110} * 100 = 1.5 * 100 = 150\%$$

Marshall

$$\boldsymbol{I_p} = \frac{\sum \boldsymbol{P_1}(\boldsymbol{Q_0} + \boldsymbol{Q_1}))}{\sum \boldsymbol{P_0}(\boldsymbol{Q_0} + \boldsymbol{Q_1})} * 100 = \frac{45(21+15)}{30(21+15)} * 100 = \frac{1620}{1080} * 100 = 1.5 * 100 = 150\%$$

Fórmula ideal de Fisher.

$$I_p = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} * \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} * 100 = \sqrt{\frac{213}{142} * \frac{165}{110}} * 100 = \sqrt{1.5 * 1.5} * 100 = \sqrt{2.25} * 100 = 1.5 * 100 = 150\%$$

$$I_q = \sqrt{\frac{\sum q_1 P_0}{\sum q_0 P_0} * \frac{\sum q_1 P_1}{\sum q_0 P_1}} * 100 = \sqrt{.79 * .79} * 100 = \sqrt{.6241} * 100 = .79 * 100 = 79\%$$

$$Iv = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} * 100 = \frac{165}{142} * 100 = 1.16 * 100 = 116\%$$

M. relativos

$$\boldsymbol{I_p} = \frac{\sum \frac{\boldsymbol{P_1}}{\boldsymbol{P_0}}}{\boldsymbol{n}} * 100 = \frac{7.5}{5} * 100 = 1.5 * 100 = 150\%$$

Mg de relativos

$$\log I_p = \log\left[\frac{\sum \frac{P_1}{P_0}}{n} * 100\right] = \log \sum \frac{P_1}{P_2} - \log n + \log 100$$

Por lo tanto log I_p = 2.1761

Su antilogaritmo = 150.0 %

Ma de relativos

$$I_p = \frac{n}{\sum \frac{P_0}{P_1}} * 100 = \frac{5}{3.35} * 100 = 1.49 * 100 = 149\% \cong 150\%$$

Tabla 9.1.1.9.2.2 Para números índices compuestos o ponderados

2001 = 100 Del precio P_1/P_o	Log de Relativos P_1/P_o	Recíproco P_o/P_1	$Q_o + Q_1$	$P_1(Q_o + Q_1)$	$P_o(Q_o + Q_1)$
1.5	0.1761	0.67	4	12	8
1.5	0.1761	0.67	5	30	20
1.5	0.1761	0.67	7	63	42
1.5	0.1761	0.67	9	108	72
1.5	0.1761	0.67	11	165	110
7.5	0.8805	3.35	36	378	252

MARSHALL

$$I_p = \frac{\sum P_1(Q_0 + Q_1)}{\sum P_0(Q_0 + Q_1)} * 100 = \frac{378}{252} * 100 = 150\%$$

También existe el índice Flores-Panse. Fue calculado por la Srita. Matemática Ana María Flores y V.G. Panse. Contiene una elaboración matemática rigurosa en el cálculo de los Q_s, lo que hace posible que el indicador (índice) resulte más apegado a la realidad económica y tenga aplicación en Paasche, Laspeyres y Fisher.

Ejemplo: para el calculo de Q_0 (consumo) su fórmula es:

$Q_0 = \sum_{i=1}^{n} c_i = \frac{\overline{«}_i N_i}{«_i} =$ Estimación del consumo por día.

Donde:

c_i = Consumo total por día en el estrato i-ésimo.

$\overline{«}_i$ = Promedio de unidades de consumo en el estrato i-ésimo, o sea convertirá total la población según su edad y sexo en unidades de consumo.

$\overline{«}_i N_i$ = Total de unidades de consumo en estrato i-ésimo.

La población se calcula tomando el sexo y la edad en unidades de consumo según la tabla de la FAO.

IX.1.1.9.3.- Pruebas matemáticas.

Se aplican para identificar qué índice es el mejor de los muchos que existen para expresar variaciones. Fisher propuso las dos que aparecen en los siguientes dos incisos. Para ello supongamos que los datos son los siguientes:

Tabla 9.1.1.9.3.1 Pruebas matemáticas.

		2001		2002	
Artículo	Unidad	P_o	Q_o	P_i	Q_i
Maíz	Bushel	2,343.00	2,679.00	0.66	3,071.00
Algodon	Libra	5,356.00	5,705.00	0.14	6,715.00
Heno	Tonelada	20,150.00	76.59	17.78	76.16
Trigo	Bushel	2.13	52.10	1.43	843.30
Avena	Bushel	0.70	1,107.00	0.46	1,444.00
Papa	Bushel	1.58	297.30	1.13	368.90
Azúcar	Libra	0.10	4,371.00	0.05	4,817.00
Cabada	Bushel	1.22	131.10	0.72	171.00
Tabaco	Libra	0.39	1,444.00	0.21	1,509.00
Linaza	Bushel	4.38	6.77	1.77	10.90

Centeno	Bushel	1.33	78.70	1.26	61.90
Arroz	Bushel	2.67	42.69	1.19	51.56

Tabla 9.1.1.9.3.2 Pruebas matemáticas.

P_0Q_0	P_1Q_0	P_0Q_1	P_1Q_1
3,597.897	1757.424	4,124.353	2,014.576
2,030.9800	792.995	2,390.540	933.385
1,543.2885	1361.7702	1,534.624	1,354.1248
2,018.9251	1364.3593	1,797.0723	1,208.4489
777.1140	504.792	1,013.688	658.464
469.7340	335.3544	582.862	416.1192
445.8420	231.663	491.334	255.301
159.2865	93.8676	207.765	122.436
563.1600	306.128	588.510	319.908
29.67291	11.9829	47.7747	19.293
104.7497	98.8472	82.3889	77.7464
113.81154	50.84379	137.6989	61.51515
11,854.4613	6,910.0274	12,998.6108	7,441.3175

IX.1.1.9. 3.1.- La prueba de reversión de factores dice:

Si se intercambian los factores P y Q en una fórmula de índice de precios (o de cantidad) de manera que se obtenga una fórmula de índices de cantidad (o de precios), el producto de los índices deberá dar el valor exacto del índice de valor:

$$\frac{P_1Q_1}{P_0Q_0}$$

Si tomamos la fórmula de Laspeyres: $\frac{\sum P_1Q_0}{\sum P_0Q_0}$ se transforma $\frac{\sum Q_1P_0}{\sum Q_0P_0}$.

esto es en un índice de cantidad, pero

$$\frac{\sum P_1Q_0}{\sum P_0Q_0}\ \frac{\sum \boldsymbol{Q_1P_0}}{\sum \boldsymbol{Q_0P_0}} \text{ es diferente de } \frac{\sum \boldsymbol{Q_1P_1}}{\sum \boldsymbol{Q_0P_0}}$$

Igualmente si tenemos la formula de Paasche :

$\frac{\sum P_1Q_1}{\sum P_0Q_1}$ se transforma en $\frac{\sum \boldsymbol{Q_1P_1}}{\sum \boldsymbol{Q_0P_1}}$; pero

$$\frac{\sum P_1Q_1}{\sum P_0Q_1} * \frac{\sum Q_1P_1}{\sum Q_0P_1} \text{ es diferente de } \frac{\sum \boldsymbol{Q_1P_1}}{\sum \boldsymbol{Q_0P_0}}$$

En cambio la formula ideal de Fisher :

$\sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} * \frac{\sum P_1Q_1}{\sum P_0Q_1}}$ al transformarse en $\sqrt{\frac{\sum \boldsymbol{Q_1P_0}}{\sum \boldsymbol{Q_0P_0}} * \frac{\sum \boldsymbol{Q_1P_1}}{\sum \boldsymbol{Q_0P_1}}}$ y multiplicarse por la anterior $\sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} * \frac{\sum P_1Q_1}{\sum P_0Q_1}} * \sqrt{\frac{\sum \boldsymbol{Q_1P_0}}{\sum \boldsymbol{Q_0P_0}} * \frac{\sum \boldsymbol{Q_1P_1}}{\sum \boldsymbol{Q_0P_1}}} = \frac{\sum \boldsymbol{P_1Q_1}}{\sum \boldsymbol{P_0Q_0}}$

Ejemplo:

$$\frac{\sum \boldsymbol{P_1Q_0}}{\sum \boldsymbol{P_0Q_0}} = 0.5824; \frac{\sum \boldsymbol{P_0Q_1}}{\sum \boldsymbol{P_0Q_0}} = \frac{12{,}968.610.8}{11{,}864.461.25} = 1.0956$$

El índice del valor: $\frac{\sum \boldsymbol{P_1Q_1}}{\sum \boldsymbol{P_0Q_0}} = \frac{7{,}441.610.8}{11{,}864.461.25} = 0.6272$

Luego en los casos de Laspeyres:

(1.0965)(0.5824) $\neq$ 0.6272; o sea que 0.6381 $\neq$ 0.6272

Con Paasche:

$$\frac{\sum \boldsymbol{P_1Q_1}}{\sum \boldsymbol{P_0Q_1}} = \frac{7,441.317.45}{12,998.610.8} = 0.5725$$

$$\frac{\sum Q_1P_1}{\sum Q_0P_1} = \frac{7,441.317.45}{6,910.027.39} = 1.0769 = 1.078868$$

tal que: (1.0769)(0.5725) $\neq 0.6272$

Trabajando con el índice ideal de Fisher:

$$\sqrt{\frac{\sum P_1Q_0}{\sum P_0Q_0} * \frac{\sum P_1Q_1}{\sum P_0Q_1}} * \sqrt{\frac{\sum Q_1P_0}{\sum Q_0P_0} * \frac{\sum Q_1P_1}{\sum Q_0P_1}} = \frac{\sum P_1Q_1}{\sum P_0Q_0}$$

Esto es: $\sqrt{(0.5824) * (0.5725)} * \sqrt{(1.0956) * (1.0769)} = 0.6272$

(0.5775)(1.0862) = 0.6272; por lo tanto 0.6272 = 0.6272

IX.1.1.9.3.2.- La prueba de reversión cronológica se expresa como sigue:

Si se intercambian los subíndices de tiempo de una fórmula de precios (o de cantidad), la fórmula resultante de precios (o de cantidad) deberá ser reciproca de la fórmula original.

Si tomamos la fórmula de Laspeyres: $\frac{\sum \boldsymbol{P_1Q_0}}{\sum \boldsymbol{P_0Q_0}}$

pero $\frac{\sum \boldsymbol{P_1Q_0}}{\sum \boldsymbol{P_0Q_0}}$ se transforma $\frac{\sum P_0Q_1}{\sum P_1Q_1}$.

$$\frac{\sum \boldsymbol{P_1Q_0}}{\sum \boldsymbol{P_0Q_0}} * \frac{\sum P_0Q_1}{\sum P_1Q_1} \neq 1.0$$

Luego no satisface la prueba; de la misma manera en el caso de Paasche:

$\frac{\sum P_1Q_1}{\sum P_0Q_1}$ se transforma $\frac{\sum P_0Q_0}{\sum P_1Q_0}$.

pero

$$\frac{\sum P_1Q_1}{\sum P_0Q_1} * \frac{\sum \boldsymbol{Q_0P_0}}{\sum \boldsymbol{Q_0P_1}} \pm 1.0$$

En cambio si aplicamos la prueba al Índice Ideal de Fisher:

$$\sqrt{\frac{\sum \boldsymbol{Q_0P_1}}{\sum \boldsymbol{Q_0P_0}} * \frac{\sum \boldsymbol{Q_1P_1}}{\sum \boldsymbol{Q_1P_0}}}$$ se cambia $$\sqrt{\frac{\sum Q_1P_0}{\sum Q_1P_1} * \frac{\sum Q_0P_0}{\sum Q_0P_1}}$$ tal que

$$\sqrt{\frac{\sum \boldsymbol{Q_0P_1}}{\sum \boldsymbol{Q_0P_0}} * \frac{\sum \boldsymbol{Q_1P_1}}{\sum \boldsymbol{Q_1P_0}}} * \sqrt{\frac{\sum Q_1P_0}{\sum Q_1P_1} * \frac{\sum Q_0P_0}{\sum Q_0P_1}} = 1.0$$

Ejemplo en el caso de Laspeyres:

$$\frac{\sum \boldsymbol{P_1Q_0}}{\sum \boldsymbol{P_0Q_0}}$$ se transforma en $$\frac{\sum P_0Q_1}{\sum P_1Q_1}$$

Recordando que : $$\frac{\sum P_0Q_1}{\sum P_1Q_1} = \frac{12{,}998.610.8}{7{,}441.317.45} = 1.7468157$$

luego (0.5824)(1.7468157) $\neq$ 1.0 porque 1.01734 $\neq$ 1.0

Con Paasche:

$$\frac{\sum \boldsymbol{P_1Q_1}}{\sum \boldsymbol{P_0Q_1}}$$ se transforma en $$\frac{\sum \boldsymbol{Q_0P_0}}{\sum \boldsymbol{Q_0P_1}}$$

donde $$\frac{\sum \boldsymbol{Q_0P_0}}{\sum \boldsymbol{Q_0P_1}} = \frac{11{,}864.461.25}{6{,}910.023.9} = 1.7169919$$

(0.5725)(1.79919) $\neq$ 1.0 es decir 0.9829778 $\neq$ 1.0

En el caso del Índice ideal de Fisher:

$$\sqrt{\frac{\sum Q_0P_1}{\sum Q_0P_0} * \frac{\sum Q_1P_1}{\sum Q_1P_0}} * \sqrt{\frac{\sum Q_1P_0}{\sum Q_1P_1} * \frac{\sum Q_0P_0}{\sum Q_0P_1}} = 1.0$$

Esto es: $\sqrt{(\mathbf{0.5824}) * (\mathbf{0.5725})} * \sqrt{(\mathbf{1.7468157}) * (\mathbf{1.71169919})} = 1.0$

(0.5774)(1.7318)=1.0 o sea que 0.99999413=1.0 por lo tanto 1.0 = 1.0

Así, puesto que sólo la fórmula propuesta por Fisher pasa estas dos pruebas matemáticas, él la llamó " Fórmula Ideal de Fisher". Cabe señalar que en México no se usa para medir variaciones ya sea de precios o de cantidades pero se incluyó este tema para que el lector vea los esfuerzos que hacen los investigadores por mejorar la metodología de variaciones de una variable.

IX.1.1.10.- ÍNDICES ESLABONADOS Y EN CADENA.

Los procesos de eslabonamientos o encadenamiento permiten hacer cambios en la muestra de bienes usados para calcular el índice ponderado compuesto.

El proceso de eslabonamiento se caracteriza por el cambio constante del año base. Por ejemplo el índice de 2000 usa como base 1999 y el de 2002 toma como base 2001.

Visto numéricamente:

Tabla 9.1.1.10.1 Índices eslabonados y en cadena.

Año	Ventas en Millones de $	Eslabón Relativo	Índice en Cadena
1998	1.5	-	136.3
1999	1.3	86.7	118.2
2000	1.1	84.6	100
2001	1.7	154.5	154.5
2002	1.9	121.1	187.09

Las limitaciones de este índice es que no se puede hacer comparaciones sobre un número determinado de años, para ello es necesario unir o encadenar los eslabones en términos de un sólo año base.

Para el año escogido como base el valor del índice es automáticamente fijado en 100, en nuestro ejemplo el año de 2000 es igual a 100. Los índices para los años siguiente a 2000 fueron determinados multiplicando el eslabón relativo de cada año por el índice en cadena del año precedente. Así, si N se refiere a un año determinado en la serie:

$$\boldsymbol{Cu} = \frac{\boldsymbol{Lu} * \boldsymbol{Cu} - 1}{100}$$

donde:

Cu = Índice de cadena del año de estudio.

Lu = Eslabón relativo.

Cu - 1 = Índice en cadena del año anterior.

Para 2002

Cu 2002 = $\frac{(121.1) * (154.5)}{100}$

Cu 2002 = 187.09

Para ir hacia atrás en el tiempo a partir de un año base la ecuación se resuelve para Cu - 1 en lugar de Cu. Así, el índice en cadena para 1998 será:

$$Cu - 1 = \frac{Cu}{Lu} * 100; Cu1998 = \frac{118.2}{86.7} * 100$$

Por tanto Cu 1998 = 136.3

Para 1999 tendremos

Cu 1999 = $\frac{100}{84.6} 100 = 118.2$

CÁLCULOS DE LOS NÚMEROS ÍNDICE[G] USADOS PARA "INFLACTAR LA INFORMACIÓN DE 1998 A 2000.

1.- Se obtuvo el índice mensual para los años de 1998 y 1999, por ser los años a que corresponden la mayoría de las empresas.

$$I_{1999} = \frac{1{,}418.20}{12} = 118.18\%; I_{1998} = \frac{1{,}199.40}{12} = 99.95\%$$

2.- Al año de 1999 o sea 118.8 se le incorporó el 30% de la inflación estimada para 2000, a fin de hacer este último igual a 100% o año base:

I_{2000} = 118.18 * 1.30 = 153.63 = 100.0%

3.- Con esta información se calcularon los números índice.

Tabla 9.1.1.10.2.Indices eslabonados y en cadena.

Año	Cálculo				Índice	
					Para dividir	Para multiplicar
2000	153.63	÷	153.63	=	1	1.0
1999	118.18	÷	153.63	=	0.77	1.3
1998	99.95	÷	153.63	=	0.65	1.54
1997	85.1	÷	153.63	=	0.55	1.82

IX.1.1.11.- Diferentes tipos de índices usados en México.

Destacan: a).- Indice Nacional de Precios al Consumidor, INPC; b).- Indice Nacional de Precios al Productor, INPP y el de la Vivienda.

[*] Indice de Precios al Consumidor, BANXICO.

Las principales diferencias (Banxico, 2002) entre el INPC y el INPP son:

Tabla 9.1.1.11.1 Diferentes tipos de índices usados en México.

INPC	INPP
Es un *indicador* (estimador porque viene de una muestra) del comportamiento de los precios de los bienes y servicios que consumen las familias en un lapso dado.	Es un *indicador* de la evolución de los precios de los bienes y servicios que forman la producción de la economía en un lapso dado.
Incluye únicamente los bienes y servicios que adquieren las familias para su consumo en un lapso dado.	Incluye: además del consumo familiar, a los bienes y servicios intermedios, de consumo del gobierno, de inversión y de exportación.
Las ponderaciones están basadas en los reportes que el INEGI levanta en los hogares, los cuales al agregarse, constituyen la Encuesta Nacional de Ingreso gasto de los Hogares, ENIGH.	Las ponderaciones se estiman con base en el Sistema de Cuentas Nacionales de México, SCNM.
Incluye las importaciones como una fracción de los bienes que consumen las familias.	No incluye a las importaciones.
Los precios son recabados en los establecimientos o fuentes de información donde las familias acuden a realizar las compras de los bienes y servicios que consumen.	Los precios se obtienen directamente de las empresas productoras de bienes o suministradoras de servicios.
Periodicidad quincenal: Los resultados se publican los días 10 y 25 de cada mes en el Diario Oficial de la Federación, en un boletín de prensa (que se emite el día anterior a su publicación en el	Periodicidad mensual. Se publica a más tardar el día 9 de cada mes en un boletín de prensa y en la hoja electrónica del Banco de México.

Diario Oficial) y en la hoja electrónica del Banco de México	
Se elabora con base en precios al consumidor final que incluyen impuestos al consumo, costos de transporte y márgenes de comercialización. Las cotizaciones son proporcionadas de manera voluntaria y se publican cada mes en el Diario Oficial de la Federación, manteniendo la confidencialidad respecto a la fuentes de información.	Los precios que se cotizan son principalmente Libre a Bordo (LAB) planta de producción. Por tanto, no incluyen impuestos al consumo, costos de transporte ni márgenes de comercialización; se proporcionan de manera voluntaria y son confidenciales.
Se calcula para 46 ciudades y a nivel nacional.	Presenta resultados a nivel nacional.

CONSTRUCCIÓN DE ÍNDICES PARA MEXICO

Ejercicio 1.

Índices relativos para un solo artículo

Referencias. La producción nacional de Maíz en grano es la siguiente:

Tabla 9.1.1.11.2 Producción de Maíz en México

Año	Producción (Ton)	PMR ($/Ton)	Valor Producción (Miles de Pesos)
1990	14,635,439.00	609.47	8,919,861.01
1991	14,251,500.00	707.31	10,080,228.47
1992	16,929,342.00	761.23	12,887,123.01
1993	18,125,263.00	767.73	13,915,308.16
1994	18,235,826.00	656.22	11,966,713.74
1995	18,352,856.00	1,091.57	20,033,427.02
1996	18,025,952.45	1,434.61	25,860,211.64
1997	17,656,258.00	1,353.75	23,902,159.27
1998	18,454,710.38	1,446.18	26,688,833.06
1999	17,706,375.63	1,454.48	25,753,569.23
2000	17,556,905.24	1,507.78	26,471,950.58
2001	20,134,312.10	1,451.07	29,216,296.26

2002	19,297,754.79	1,500.56	28,957,438.93
2003	20,701,420.03	1,618.01	33,495,104.62
2004	21,685,833.34	1,678.59	36,401,622.99
2005	19,338,712.89	1,577.93	30,515,135.23
2006	21,893,209.25	2,010.55	44,017,391.86
2007	23,512,751.85	2,441.99	57,417,904.89
2008	24,410,278.53	2,817.04	68,764,731.03
2009	20,142,815.76	2,802.05	56,441,176.90
2010	23,301,878.98	2,816.48	65,629,276.11

Fuente: Elaboración propia con datos de SAGARPA.

Considerando 1990 = 100, es decir, año base.

Basándonos en 1990 como 100 decimos que 14, 635,439 Toneladas será = 100 y con esta referencia obtendremos el índice para cada año.

Ejemplo: Para 1993 = 18, 125,263 Toneladas basándonos en el año 1990 = 100

$$1993 = \left(\frac{18{,}125{,}263}{14{,}635{,}439}\right)100 = 123.845$$

Ejercicio 2

Tabla 9.1.1.11.3 Índices Relativos para Precio, Cantidad y Valor de la Producción de Maíz

Año	Producción (Ton)		PMR ($/Ton)		Valor Producción (Miles de Pesos)	
	Unidades	Índice	Precio	Índice	Valor	Índice
1990	14,635,439.00	100	609.47	100	8,919,861.01	100
1991	14,251,500.00	97.38	707.31	116.053292	10,080,228.47	113.008807
1992	16,929,342.00	115.67	761.23	124.900323	12,887,123.01	144.476724
1993	18,125,263.00	123.85	767.73	125.966824	13,915,308.16	156.003643
1994	18,235,826.00	124.6	656.22	107.670599	11,966,713.74	134.158074
1995	18,352,856.00	125.4	1,091.57	179.101514	20,033,427.02	224.593489
1996	18,025,952.45	123.17	1,434.61	235.386483	25,860,211.64	289.917204
1997	17,656,258.00	120.64	1,353.75	222.119218	23,902,159.27	267.965602
1998	18,454,710.38	126.1	1,446.18	237.284854		299.206827

					26,688,833.06	
1999	17,706,375.63	120.98	1,454.48	238.646693	25,753,569.23	288.721643
2000	17,556,905.24	119.96	1,507.78	247.391996	26,471,950.58	296.775371
2001	20,134,312.10	137.57	1,451.07	238.087191	29,216,296.26	327.542057
2002	19,297,754.79	131.86	1,500.56	246.20736	28,957,438.93	324.640024
2003	20,701,420.03	141.45	1,618.01	265.478202	33,495,104.62	375.511508
2004	21,685,833.34	148.17	1,678.59	275.417986	36,401,622.99	408.096303
2005	19,338,712.89	132.14	1,577.93	258.901997	30,515,135.23	342.103259
2006	21,893,209.25	149.59	2,010.55	329.884982	44,017,391.86	493.476208
2007	23,512,751.85	160.66	2,441.99	400.674356	57,417,904.89	643.708516
2008	24,410,278.53	166.79	2,817.04	462.21143	68,764,731.03	770.917069
2009	20,142,815.76	137.63	2,802.05	459.751916	56,441,176.90	632.75848
2010	23,301,878.98	159.22	2,816.48	462.119546	65,629,276.11	735.765681

Fuente: Elaboración propia con datos de SAGARPA.

Si ahora cambiamos de base digamos 2000 = 100, haciendo los cálculos para los precios tendremos:

$$1990 = \left(\frac{100}{119.96}\right)100 = 83.36$$

Siendo 100 la base anterior ya que considerábamos a 1990 como 100 ahora este lo dividiremos entre el valor de la nueva base que será 2000 = 119.96 en índice de acuerdo a las toneladas y a la base anterior y obtenemos el nuevo índice.

Tabla 9.1.1.11.4 Cambio de base en la producción de Maíz

Año	**Producción (Ton)**		
	Unidades	**Índice base 1990**	**Índice base 2000**
1990	14,635,439.00	100.00	83.36
1991	14,251,500.00	97.38	81.17
1992	16,929,342.00	115.67	96.43
1993	18,125,263.00	123.85	103.24
1994	18,235,826.00	124.60	103.87

1995	18,352,856.00	125.40	104.53
1996	18,025,952.45	123.17	102.67
1997	17,656,258.00	120.64	100.57
1998	18,454,710.38	126.10	105.12
1999	17,706,375.63	120.98	100.85
2000	17,556,905.24	119.96	100.00
2001	20,134,312.10	137.57	114.68
2002	19,297,754.79	131.86	109.92
2003	20,701,420.03	141.45	117.91
2004	21,685,833.34	148.17	123.52
2005	19,338,712.89	132.14	110.15
2006	21,893,209.25	149.59	124.70
2007	23,512,751.85	160.66	133.92
2008	24,410,278.53	166.79	139.04
2009	20,142,815.76	137.63	114.73
2010	23,301,878.98	159.22	132.72

Fuente: Elaboración propia con datos de SAGARPA.

Lo mismo puede hacerse para los Precios y el Valor de la Produccion como anteriormente se elaboro para índices en general en el Cuadro 2.

Aplicaciones para deflactar e inflactar.

Como antes se indicó, la deflactación se hace lo mismo para una serie cronológica como para el análisis comparativo en dos años de un fenómeno en términos reales.

Podemos definir la inflación como el aumento sustancial y sostenido del nivel general de precios. Detrás de este fenómeno están la cantidad total del dinero en la economía y la lucha de los distintos agentes económicos por el reparto de la renta.

Por ello para evaluar estos cambios se necesitan actualizar el valor de los bienes e ingresos.

La técnica que se utiliza es la deflactación de valores corrientes transformándolos en valores constantes a través de la aplicación de un índice como pueden ser el Índice Nacional de Precios al Consumidor (INPC).

El Indice Nacional de Precios al Consumidor es un indicador económico que se emplea recurrentemente, cuya finalidad es la de medir a través del tiempo la variación de los precios de una canasta fija de bienes y servicios representativa del consumo de los hogares. El INPC es el instrumento estadístico por medio del cual se mide el fenómeno económico que se conoce como inflación. Así, el INPC es la medida de la inflación por explicar una similitud.

Tabla 9.1.1.11.5. Deflactación

Periodo	Salario Nominal	INPC	Salario Real
2001	37.57	95.424	39.372
2002	39.74	100.224	39.651
2003	41.53	104.782	39.635
2004	43.29	109.694	39.464
2005	45.24	114.069	39.660
2006	47.05	118.209	39.802
2007	48.88	122.898	39.773
2008	50.84	129.197	39.351
2009	53.19	134.071	39.673

Fuente: INEGI, con base en cifras de la Comisión Nacional de Salarios Mínimos.

En el cuadro numero 4 se obtienen los salarios reales con ayuda del Índice Nacional de Precios al Consumidor, teniendo estos datos la deflactacion consiste en basarnos en el índice de precios para asi obtener el salario real como e muestra en el siguiente ejemplo:

Ejemplo para 2004:

Como el INPC es de 109.694, sabemos que es un índice por lo que su valor esta multiplicado por 100. Así que tomaremos el valor de variación que será 1.0969. De ese modo tendremos lo siguiente:

$$2004 = \frac{43.29}{1.0969} = 39.46$$

Existen dos maneras de realizarlo, la primera es dividiendo

Ejemplo 2010/01

El INPC de este periodo es 96.58 y si tomamos como base 2012/02 = 104.5

$$2010/01 = \frac{96.58}{104.5} = 0.92$$

Multiplicando

$$2010/01 = \frac{104.5}{96.58} = 1.082$$

Los resultados para cada periodo se pueden observar en el Cuadro 5 Inflactacion

En el siguiente cuadro tenemos al igual que en el cuadro 4 el INPC pero ahora lo utilizaremos de 2 formas: multiplicar y dividir, que se explico su procedimiento anteriormente, de este modo los mostramos en una tabla para que el lector pueda resolverlos y comprobar su aprendizaje.

Tabla 9.1.1.11.6. Inflactacion

Periodo	ÍNPC	Para dividir	Para Multiplicar
2010/01	96.58	0.924	1.082
2010/02	97.13	0.930	1.076
2010/03	97.82	0.936	1.068
2010/04	97.51	0.933	1.072
2010/05	96.90	0.927	1.078
2010/06	96.87	0.927	1.079
2010/07	97.08	0.929	1.076
2010/08	97.35	0.932	1.073
2010/09	97.86	0.936	1.068
2010/10	98.46	0.942	1.061
2010/11	99.25	0.950	1.053
2010/12	99.74	0.955	1.048
2011/01	100.23	0.959	1.043
2011/02	100.60	0.963	1.039
2011/03	100.80	0.965	1.037
2011/04	100.79	0.965	1.037
2011/05	100.05	0.957	1.044
2011/06	100.04	0.957	1.045
2011/07	100.52	0.962	1.040
2011/08	100.68	0.963	1.038
2011/09	100.93	0.966	1.035
2011/10	101.61	0.972	1.028

2011/11	102.71	0.983	1.017
2011/12	103.55	0.991	1.009
2012/01	104.28	0.998	1.002
2012/02	104.50	1	1

Fuente: Elaboración propia con datos de INEGI

Ejemplo $100 millones de ventas de 2010/1, 2010/2 y 2010/3 equivalen a precios de 2011 a:

Tabla 9.1.1.11.7. Ventas

Año	Ventas		(Millones de pesos de cada empresa)			
2010/1	$100 ÷	0.92 =	$ 108 mil	=	$100 *	1.08
2010/2	$100 ÷	0.93 =	$ 107 mil	=	$100 *	1.07
2010/3	$100 ÷	0.936 =	$ 106 mil	=	$100 *	1.06

Fuente: Elaboración propia.

Cálculo de la inflación mensual acumulada.

Tabla 9.1.1.11.8 Inflación acumulada

Periodo	Índice nacional de precios al consumidor	Variación porcentual	Base para aplicar la inflación	Inflación mensual en %	Inflación acumulada en %
2011/01	100.228	3.782	100.0	3.782	3.782
2011/02	100.604	3.572	103.782	3.707437	7.489437
2011/03	100.797	3.040	107.489	3.267149	10.75659
2011/04	100.789	3.361	110.757	3.722161	14.47875
2011/05	100.046	3.249	114.479	3.719745	18.19849
2011/06	100.041	3.276	118.198	3.872736	22.07123
2011/07	100.521	3.547	122.071	4.330065	26.40129
2011/08	100.68	3.424	126.401	4.32759	30.72888
2011/09	100.927	3.137	130.729	4.10067	34.82955

2011/10	101.608	3.200	134.830	4.314546	39.1441
2011/11	102.707	3.480	139.144	4.842215	43.98631
2011/12	103.551	3.820	143.986	5.500277	49.48659
Suma		40.888		49.48659	

(Base segunda quincena de diciembre 2010=100)

Fuente: Elaboración propia con datos de INEGI

Para obtener la tasa mensual acumulada, no se debe sumar las tasas de inflación de cada mes, se debe multiplicar y después sumar; para así acumular correctamente las tasas de inflación de cada mes.

Así al empezar el mes de enero de 2011, se parte de la base 100 (columna I renglón A). La tasa de inflación del mes de enero fue de 3.78% luego la tasa de inflación acumulada al final del mes fue del 3.78 (columna IV renglón A)

El cálculo de la inflación acumulada al 29 de febrero es :

103.78 x 0.0352 = 3.7074 + 3.78 = 7.48 % (columna IV renglón B).

Ejemplos adicionales.

PARA NÚMEROS ÍNDICES:

Recordamos que el índice de Laspeyres nos sirve para medir el costo total en cualquier año de comparación de un conjunto fijo de artículos comprados en el año base y el índice de Paasche nos sirve para medir el costo total de los artículos en el año de comparación relativos al costo que podrán haber sido comprados si se hubiera adquirido en el año base.

Laspeyres.

Método del año base $I_p = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} * 100$

Paasche

$$I_p = I_p = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} * 100$$

Método del año de comparación

Marshall

$$I_p = \frac{\sum P_1 (Q_0 + Q_1)}{\sum P_0 (Q_0 + Q_1)} * 100$$

$$I_p = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} * \frac{\sum Q_1 P_1}{\sum Q_0 P_1}} * 100$$

$$I_q = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} * \frac{\sum Q_1 P_1}{\sum Q_0 P_1}} * 100$$

$$Iv = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} * 100$$

donde Iq = índice de cantidad

Iv = índice de valor

Cálculo de cada uno de los números índice:

Laspeyres.

$$I_p = \frac{\sum P_1 Q_0}{\sum P_0 Q_0} * 100 = \frac{3,881,315,511.1}{2,737,021,850.1} * 100 = 141.80\%$$

Paasche

$$I_p = \frac{\sum P_1 Q_1}{\sum P_0 Q_1} * 100 = \frac{5,379,469,107.0}{3,886,022,168.7} * 100 = 138.4\%$$

Marshall

$$I_p = \frac{\sum P_1 (Q_0 + Q_{1)})}{\sum P_0 (Q_0 + Q_1)} * 100 = \frac{74,027(2,625,975 + 2,762,822)}{70,574(2,625,975 + 2,762,822)} * 100 = 104.89\%$$

Fórmula ideal de Fisher.

$$I_p = \sqrt{\frac{\sum P_1 Q_0}{\sum P_0 Q_0} * \frac{\sum P_1 Q_1}{\sum P_0 Q_1}} * 100 = \sqrt{\frac{3,881,315,511.1}{2,737,021,850.1} * \frac{5,379,469,107.0}{3,886,022,168.7}} * 100 = 140.089\%$$

$$I_q = \sqrt{\frac{\sum q_1 P_0}{\sum q_0 P_0} * \frac{\sum q_1 P_1}{\sum q_0 P_1}} * 100 = \sqrt{\frac{3,886,022,168.7}{2,737,021,850.1} * \frac{5,379,469,107.0}{3,881,315,511.1}} * 100 = 140.27\%$$

$$Iv = \frac{\sum P_1 Q_1}{\sum P_0 Q_0} * 100 = \frac{5,379,469,107.0}{2,737,021,850.1} * 100 = 196.54\%$$

Tabla 9.1.1.11.9 Cálculo de números índice

Cultivo	2009		2010		P1Q0	P0Q0	P1Q1	P0Q1
	PMR ($/Ton) P0	Producción (Ton) Q0	PMR ($/Ton) P1	Producción (Ton) Q1				
ACELGA	2,937	6,890	3,712	7,702	25,577,019.9	20,235,926.4	28,593,056.1	22,622,142.1
AGAPANDO (Gruesa)	224	8,000	143	4,347	1,142,000.0	1,795,600.0	620,505.7	975,639.3
AJO	7,397	56,088	15,104	47,429	847,182,275.5	414,868,858.6	716,390,490.7	350,819,550.7
AJONJOLI	12,044	28,523	11,910	37,289	339,711,960.6	343,531,253.2	444,105,656.5	449,098,620.1
ALBAHACA	10,034	2,553	12,260	3,723	31,295,529.2	25,611,466.4	45,650,986.4	37,359,607.9
ALBRICIA	1,277	1,232	1,656	1,241	2,040,426.1	1,572,795.8	2,054,934.3	1,583,979.0
ALCACHOFA	11,167	215	6,717	686	1,444,103.4	2,400,999.6	4,607,697.4	7,660,863.8
ALGODON HUESO	6,743	278,526	9,305	440,489	2,591,666,967.8	1,877,967,348.6	4,098,732,028.6	2,970,013,129.0
ALHELI	1,808	2,064	1,645	1,968	3,396,146.9	3,731,629.4	3,238,811.8	3,558,752.3
ALHELI (Gruesa)	50	1,900	60	1,750	114,000.0	95,000.0	105,000.0	87,500.0

ALHELI (Manojo)	16	258,580	16	353,750	4,028,676.4	4,134,694.2	5,511,425.0	5,656,462.5
ALMACIGO (Planta)	2	1,975,000	2	1,856,400	3,930,250.0	3,950,000.0	3,694,236.0	3,712,800.0
ALPISTE	8,825	112	4,981	197	555,475.9	984,103.6	982,132.8	1,739,986.3
ALPISTE ORNAMENTAL	11	1,800	15	1,980	27,000.0	19,800.0	29,700.0	21,780.0
AMARANTO	8,039	4,493	6,499	3,870	29,203,679.4	36,122,374.1	25,152,445.8	31,111,355.7
SUMA	70,574	2,625,975	74,027	2,762,822	3,881,315,511.1	2,737,021,850.1	5,379,469,107.0	3,886,022,168.7

Fuente: Elaboración propia con datos de SAGARPA

M. relativos

$$I_p = \frac{\sum \frac{P_1}{P_0}}{n} * 100 = \frac{16.2}{15} * 100 = 1.08 * 100 = 108\%$$

Mg de relativos

$$\log \boldsymbol{I_p} = \log \left[\frac{\sum \frac{\boldsymbol{P_1}}{\boldsymbol{P_0}}}{\boldsymbol{n}} * 100 \right] = \log \sum \frac{\boldsymbol{P_1}}{\boldsymbol{P_2}} - \log \boldsymbol{n} + \log 100$$

Por lo tanto log I_p = 2.0334

Su antilogaritmo = 108 %

Ma de relativos

$$I_p = \frac{n}{\sum \frac{P_0}{P_1}} * 100 = \frac{15}{15.5} * 100 = 96.77\%$$

MARSHALL

$$I_p = \frac{\sum P_1 (Q_0 + Q_1)}{\sum P_0 (Q_0 + Q_1)} * 100 = \frac{9{,}260{,}784{,}618.2}{6{,}623{,}044{,}018.9} * 100 = 139.82\%$$

Tabla 9.1.1.11.9 Construcción de Índices

Cultivo	P1/P0	P0/P1	Q0 + Q1	P0(Q0 + Q1)	P1(Q0 + Q1)
ACELGA	1.26394114	0.79117608	14,592	42858068.52	54170076
AGAPANDO (Gruesa)	0.63599911	1.57232925	12,347	2771239.26	1762505.7
AJO	2.04204837	0.48970437	103,517	765688409.3	1563572766
AJONJOLI	0.98888226	1.01124274	65,812	792629873.4	783817617.1
ALBAHACA	1.2219343	0.8183746	6,276	62971074.27	76946515.61
ALBRICIA	1.29732418	0.77081736	2,473	3156774.871	4095360.384
ALCACHOFA	0.60145924	1.66262305	901	10061863.44	6051800.76
ALGODON HUESO	1.38003835	0.72461754	719,015	4847980478	6690398996
ALHELI	0.91009757	1.09878329	4,032	7290381.745	6634958.7
ALHELI	1.2	0.83333333	3,650	182500	219000

(Gruesa)					
ALHELI (Manojo)	0.97435897	1.02631579	612,330	9791156.7	9540101.4
ALMACIGO (Planta)	0.995	1.00502513	3,831,400	7662800	7624486
ALPISTE	0.5644486	1.7716405	309	2724089.918	1537608.738
ALPISTE ORNAMENTAL	1.36363636	0.73333333	3,780	41580	56700
AMARANTO	0.80846512	1.23691175	8,363	67233729.84	54356125.2
SUMA	16.2	15.5	5,388,797	6,623,044,018.9	9,260,784,618.2

Fuente: Elaboración propia con datos de SAGARPA

CAPITULO X.- SERIES DE TIEMPO: SUS CARACTERÍSTICAS

Este tema se empezó a estudiar con rigor técnico y profusión durante la década de los 70 del siglo pasado, ello se debió a que la metodología de la econometría tradicional o clásica empezó a fallar, despertar desconfianza el su uso como instrumento para hacer análisis económico, principalmente para hacer estudios de las variables relevantes macro (planeación, análisis de estructura) y micro (valores de las acciones de una empresa en el mercado de valores). Donde mostró su ineficiente significativamente fue en los estudios de predicción en virtud que las crisis petrolera alteró mucho los precios de este energético, como el abandono de la paridad de 35 dólares con 1 onza de oro, al igual que la deuda externa que no pudieron pagar algunos países como México, principalmente, trastocaron las variables de los sistemas económicos del mundo, de manera que ya no resultaba confiable estimar los valores de una variable dependiente en función de los valores de varias variables independientes: análisis multivariable.

Ello orilló a los investigadores a buscar nuevos métodos de estimación; dentro de las opciones destaca la metodología de Box &Jenkins, la cual se caracteriza por estimar los valores futuros de la variable dependiente a partir de sus valores del pasado; es decir depender de sus valores históricamente observados, que actúan como variable independiente en el uso de MCO para hacer estimaciones y predicciones, llamado por esa razón análisis univariable, esencia de lo que se conoce como econometría moderna o de series de tiempo.

Box & Jenkins son los autores de este enfoque metodológico. sus modelos se caracterizan porque se tiene en cuenta la dependencia existente entre los datos de la misma variable en momentos diferentes de tiempo. Cada dato en un momento dado es modelado en función de los valores anteriores; sus análisis se basan en el modelo ARIMA, mismo que describe un $Y_t = Y_{t-1} + U_t$ como una función lineal de sus valores anteriores (Y_{t-1}) y errores debidos al azar (aleatorios, U_t); debe agregarse que también puede incluir un componente cíclico o estacional. Con este análisis de los datos se logra obtener un modelo "sin necesidad de ninguna otra información de variables auxiliares o relacionadas" (Pérez, 2007:547). Para realizar este análisis recomiendan usar tamaños de muestras (n) mayores a 50 datos de Y_t. Así, *modelar los valores* de esta variable en el tiempo es *deriva*r un modelo ARIMA que se ajuste al conjunto de dichos valores de Y_t. Lo anterior hizo famosa su frase de que "la información hable por si misma", que en otras palabras significa según mi interpretación, que de acuerdo con la observación de su variabilidad es que se determina la estructura matemática, ecuaciones, que habrán de servir para predecir Y_t. Así, dado que cada observación está en función de valores suyos en el pasado, es que también por eso se le conoce como análisis univariado.

X.1.- ¿Qué es, cómo se define una serie de tiempo univariante?

Una serie temporal se define como el registro de una sucesión de puntos o valores en el tiempo (día, mes, año) de una variable económica de interés para el investigador, i.e., sus valores en el tiempo pueden moverse con ciertas tendencia identificadas o con oscilaciones prolongadas, oscilaciones que no siempre son regulares. La relativa irregularidad con que se mueven hace poco apropiado su análisis con los procedimientos estadísticos estándar o de econometría tradicional, en virtud de que éstos se basan en los **supuestos** de que los datos tienen una propiedad conocida como ESTACIONARIEDAD, misma que muchas series no la tienen y a las que se les puede denominar INTEGRADAS que en opinión de Granger (2003) también se les llama incorrectamente: SERIES NO ESTACIONARIAS.

Un modelo univariante predice valores futuros de una serie de tiempo con base en valores anteriores de la misma serie de tiempo. Se usa un modelo univariable para identificar un patrón de datos, con el objetivo de que este padrón se repite en el futuro el valor se extrapola con el objetivo de generar predicciones.

¿De conformidad con las citadas fluctuaciones, de cuantas maneras podemos clasificar o describir a las series temporales?

De acuerdo con Daniel Peña (2005) las series univariantes pueden ser:

1.- Series estables o estacionarias, que se caracterizan por que sus datos fluctúan en torno de un nivel constante, que puede ser la media aritmética o la dispersión mínima de los mismos con respecto a ella y que se mide con la varianza, situación que muestra que no hay tendencias crecientes o decrecientes, lo cual es bueno porque indica que hay poca dispersión o alejamiento de sus datos del citado nivel.

2.- Series no estables o no estacionarias, éstas son aquellas cuyos datos no fluctúan alrededor de un nivel constante sino que éstos no son claramente estables en el tiempo, además de que tienen un comportamiento o tendencia que se acrecienta con el tiempo y que puede ser positiva o negativa.

3.- Las Series estacionales o no estables pueden comprender fluctuaciones como las descritas en 1 y 2 y además, un “comportamiento superpuesto que se repite periódicamente”. En otras palabras. Estos datos muestran tendencias definidas como estacionalidades, lo cual plantea problemas de cálculos precisos y de interpretación ya que si por ejemplo, calculamos la media aritmética de los datos con tendencia y de los estacionales, sus valores serán diferentes. En síntesis estas series se caracterizan por contener niveles variables : tendencias crecientes o decrecientes combinadas con otros datos de estabilidad estacional.

Conviene agregar que el “comportamiento superpuesto y que se repite periódicamente” puede ser de varios tipos: tendencia secular, cíclico, estacional e irregular.

Así, a la situación en que la media aritmética depende o está en función del periodo que se escoja de la serie temporal, se le conoce como ESTACIONALIDAD, que es diferente la de la ESTACIONARIEDAD antes definida.

X.1.1.- Evaluación.

Así, luego entonces para propósitos de predicción es bueno que la serie temporal sea estacionaria y es malo que muestre estacionalidad por las razones antes señaladas, de manera que esta última debe de eliminarse para pronosticar con mayor certeza.

Ejemplos de estos tres tipos de series son los que se describen en la siguiente sección de ejercicios con Eviews.

Ahora bien, para complementar la exposición del tema conviene preguntar:

¿Qué es, cómo se definen las series de tiempo multivariantes?

Decimos que cuando se estudia el comportamiento de los datos de varias series en el tiempo a la vez, se entiende que es un estudio de series multivariantes porque se determina su relación dinámica entre ellas. Dicha relación puede ser en dos direcciones, es decir, puede ser que mientras los valores de una serie aumenten, los de la otra disminuyan o viceversa. También, que ambos aumenten o disminuyan en la misma dirección. Esto último es el umbral de los estudios que se hace con la metodología de cointegración que se expondrá más adelante.

La aplicación de modelos causales requiere de la identificación de otras variables que se relacionan con la variable que se quiere pronosticar, de esta forma, se desarrolla un modelo estadístico que evalúa la relación de las variables.

X.1.2.- ¿Cómo se identifican la estacionalidad y la estacionariedad de las series temporales?

En el caso de la **estacionalidad**, en virtud de que en los siguientes ejercicios # 1 y 2 se exponen tanto los métodos para identificar como para eliminar el "comportamiento superpuesto y que se repite periódicamente", aquí simplemente diremos que suelen usarse los métodos clásicos de tendencias deterministas, que tienen como fundamento los ´métodos de regresión hecha con MCO. Destacan el gráfico, el del alisamiento de Holt, Brown, Winthers, de medias móviles, etc,

X.2.-ESTACIONALIDAD: Conceptos básicos para su identificación

Como antes se indicó, dentro del tipo de datos que más se usan para hacer análisis económico destacan los de las series de tiempo, Y_t, cada una de ellas se define como una sucesión de puntos u observaciones en el tiempo de una(s) variable(s) económica(s) en estudio, ejemplo: PIB, consumo, inversión, tipo de cambio, inflación, etc.

Sus datos se usan más en análisis macroeconómicos y, en contraposición, los de corte transversal en los microeconómicos. Son más difíciles de analizar porque sus datos por lo general están relacionados con sus valores anteriores, es decir, aun cuando se acepta lo anterior, casi nunca suponemos que son independientes en el tiempo.

Lo anterior es muy importante porque por ejemplo el valor del índice inflacionario del mes anterior definitivamente será muy parecido al de este mes, es decir, "influye" en el valor actual, luego *no* hay mucha "independencia" temporal entre ellos.

Al igual que la característica antes descrita, los datos de series de tiempo tienen otra característica: la de la periodicidad con que se obtienen, que puede ser cada 10 años como los de población, anualmente como los del PIB, mensual como la inflación, etc. Ello puede revelar "características estacionales" que influyen en la determinación de su "tendencia temporal".

Debido a lo anterior es que conviene analizar su distribución o *tendencia* en el tiempo con objeto de obtener esta última "libre de influencias transitorias o estacionales".

X.2.1-Tipos de componentes o movimientos de las series, Yt.

La teoría de series de tiempo establece que éstas, Y_t, pueden presentar cuatro tipos de oscilaciones, ya sea en forma independiente o simultáneamente, ellas son: 1. Tendencia; 2. Variaciones estacionales; 3. Variaciones cíclicas; y 4. Variaciones residuales o aleatorias.

X.2.1.1.- Identificación

La *tendencia secular*, T_t, se identifica como un movimiento de Y_t en el largo plazo. Por su parte las *variaciones estacionales*, E_t, son movimientos observados en periodos iguales de tiempo (diarios, semanales, meses, trimestres, etc.) en el año, es decir, que se repiten o replican sucesivamente, a veces en varios años. Las *variaciones cíclicas*, C_t, son movimientos *similares* que se observan en periodos mayores a un año, es decir, son oscilaciones en el largo plazo. Finalmente, las *variaciones residuales o irregulares* (como les llaman otros autores), R_t, son oscilaciones que muestran un comportamiento irregular o difícilmente reconocibles de Y_t, que se gestan en condiciones circunstanciales (guerras, epidemias, etc.) no permanentes.

Por la importancia que tienen las variaciones irregulares como fenómenos aislados, insólitos, a veces coyunturales, así como por el impacto o incidencia que tienen en la metodología instrumentada para determinar la bondad de ajuste de ésta para pronosticar con mayor certeza valores futuros de Yt, se considera pertinente abundar sobre las mismas diciendo que por su característica circunstancial, algunos estudiosos del tema (Mason et al, 2001:650) consideran necesario dividirlas en dos grupos: el que agrupa las fluctuaciones *episódicas* y el que agrupa las que se derivan de éstas: las *residuales*.

Ellos consideran que las episódicas aun cuando se pueden identificar no se pueden predecir puesto que por ejemplo, un incendio o una pandemia, es observable y por consiguiente se puede registrar pero no es posible pronosticarlo(a) en el futuro. Profundizando, como sucedió con la pandemia de "gripa" en la Ciudad de México en el año 2008", en virtud de que no se eliminó totalmente, una vez que se erradicó esta variación episódica, a los *restos* del efecto que causó en el quebranto de la salud, generalmente *desconocidos* (dada su periodicidad o intensidad), se les acostumbra llamar oscilaciones o *variaciones residuales* que, como se observa, por su naturaleza, son difíciles de identificar de inmediato y por consiguiente, también son impredecibles.

X.2.1.2.-Nuevo escenario de análisis

Esta última connotación <u>abre el escenario del estudio concienzudo de los métodos de predicción </u>que para ser aceptados como apropiados deben considerar o incluir también el manejo de factores desconocidos visualmente, porque estamos conscientes de que existen en la realidad aunque no los veamos.

En la terminología de series de tiempo a estos factores desconocidos se les denomina "*variaciones aleatorias*" que por su importancia, reiteramos que no son identificables fácilmente y por ende, no pueden pronosticarse. Mason, et al, deriva de esta situación el siguiente corolario categórico: "Por supuesto, ninguna variación, sea episódica o residual, puede proyectarse al futuro" (2001:650).

Esta afirmación suscita dudas como se verá más adelante; por el momento aceptemos que efectivamente existen factores desconocidos que deben ser incluidos en alguna parte del algoritmo que se use para pronosticar, también aceptemos en darles la connotación aleatoria.

Al respecto, al identificarlos como variable aleatoria, recordemos que sus valores constituyen lo que se conoce como una distribución probabilística cuya característica es que al igual que una distribución de frecuencias, describe los resultados de un experimento, pero en lugar de mostrar la frecuencia con que se repiten, muestra su probabilidad de ocurrencia. En este contexto, si tomamos el experimento, por ejemplo del lanzamiento de un dado una vez, sabemos que apriorísticamente *no conocemos* los resultados posibles (luego les podemos llamar aleatorios) de ese lanzamiento del dado; sin embargo, si recordamos que con el método clásico o teórico del cálculo de la probabilidad de un evento de interés para el investigador, éste se puede identificar sin llevar a cabo el experimento porque al abstraernos en la teoría de la probabilidad sabemos que al lanzar el dado hay seis resultados posibles: 1,2,3,4,5 y 6; también por el método de cálculo familiar para todos nosotros, sabemos que un sexto es la probabilidad de ocurrencia de cada una de las caras del dado.

Vemos que con el uso de la teoría de la probabilidad si es posible identificar un evento de una variable aleatoria como le estamos llamando, otros autores le llaman “resultado posible” de un experimento. Creo que esta opinión es congruente con la percepción que tiene Wooldridge (2009: 847) de variable aleatoria cuando dice que ésta tiene un *resultado incierto*. Yo coincido en que es incierto porque es desconocido pero le agrego que con el uso de la probabilidad que es identificable o imaginable.

Al conocer la probabilidad de ocurrencia de cada uno de los eventos desconocidos, constituidos por las caras de un dado, como antes se indicó ello da origen a una distribución probabilística. En este contexto, si ahora a cada resultado se le llama variable aleatoria, y si esta corresponde a una serie de tiempo, entonces cada uno de ellos está indexado temporalmente de manera que puede verse como un “proceso”, entonces estamos dando vida a lo que se conoce como *“procesos estocásticos”,* que es sinónimo de una serie temporal aleatoria(Wooldridge, 2009: 341) . Una conclusión que sacamos de estos razonamientos es que con la variable aleatoria se constituyen distribuciones probabilísticas, que ahora también llamaremos “procesos” estocásticos”, sobre los cuales profundizaremos más adelante.

Sin adentrarnos en el debate continuemos *usándola* como la expresión de lo desconocido y llevémosla al campo de la econometría, concretamente al uso del método de MCO, Mínimos Cuadrados Ordinarios, para hacer estimaciones de los parámetros de los modelos MLS, Modelo Lineal Simple, y MLM, Modelo Lineal Múltiple. Al respecto, muchos honorables autores lo vienen diciendo desde hace tiempo pero simplemente por ser reciente la obra de Wooldrige (2009) ésta se utilizará como referencia teórica respetable para mostrar el uso de la variable aleatoria como sinónimo del “término de error o perturbación”, que denominaremos con la letra e, que se usa en la ecuación de regresión para identificar que además de la(s) variable(s) Explicativa(s) de Yt, digamos Xt, se está consciente de que también, aun cuando no se conozcan, existen otras que inciden en el valor estimado de Y. Por ello es que *al término de error también se le llama variable aleatoria*, independientemente de que también se diga que representa los errores de medición de las variables usadas en la regresión o, que se refiere a “factores no observables” (Woolddridge, 2009:23) . En resumen, esta variable aleatoria tiene una distribución probabilística ahora llamada proceso estocástico, que tiene ciertas características estadísticas: media cero y varianza determinada, es decir, tiene una distribución normal.

Generalizando, si se toma en cuenta que con la base de datos de la variable Y_t que se usa para hacer las estimaciones con MCO, entonces sus datos o variables aleatorias constituyen un proceso estocástico o proceso de series de tiempo, i.e., estocasticidad es sinónimo de aleatoriedad.

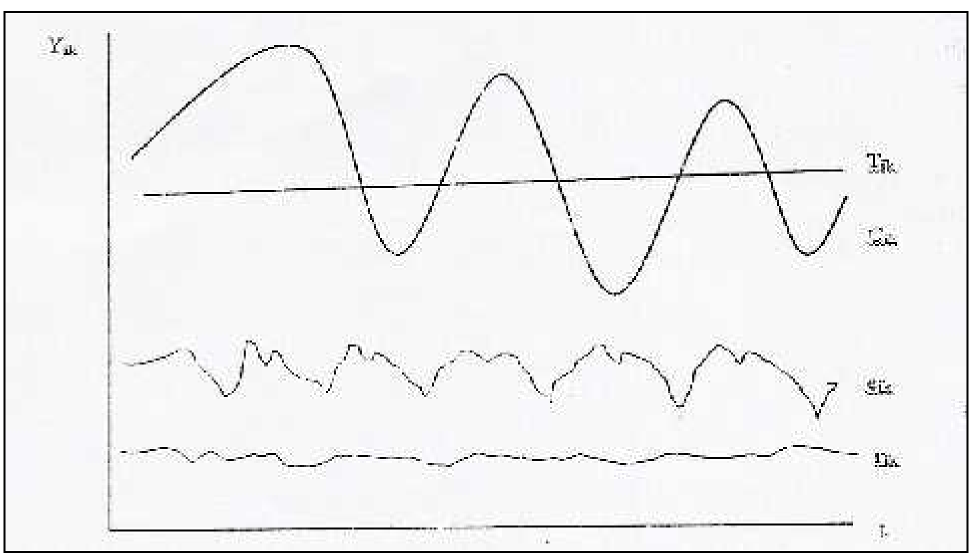
Así, para estudiar una serie de tiempo entendida como un proceso estocástico, en un periodo de tiempo determinado, donde cada uno de sus puntos es una variable aleatoria y por ello en su conjunto constituyen un proceso aleatorio, **es necesario** que se estudie ya sea en sus niveles (valores originales) o en sus variaciones o cambios (tasas) de la variable o serie de tiempo, tanto gráfica como numéricamente.

En conclusión, esperando no haber provocado una digresión sino por el contrario, haber destacado el origen, uso y alcance de las fluctuaciones residuales, así como la razón por las cuales se les llama aleatorias y el por qué constituyen los elementos de una distribución llamada probabilística o proceso estocástico, *continuemos con la explicación general de los cuatro movimientos u oscilaciones que puede tener Y_t en el tiempo.*

X.2.1.3.-Representación gráfica

Luego entonces, diremos que gráficamente los cuatro tipos de oscilaciones que puede tiene Yt, antes descritas, se manifiestan así:

Gráfico 10.2.1.3.1-Representación gráfica



Cabe señalar que estos cuatro tipos de movimientos de las series históricas suelen combinarse, es decir, sus movimientos pueden mezclarse en el periodo de tiempo de análisis de Y_t, propiciando así el surgimiento de una “nueva serie”.

Por ejemplo puede gestarse la mezcla:

a) Aditivo: $Y_t = T_t + C_t + E_t + R_t$;
b) Multiplicativo: $Y_t = T_t*C_t*E_t*R_t$; y
c) Mixto: $Y_t = T_t + C_t*E_t + R_t$

X.2.1.3.1.- Identificación del tipo de variaciones que muestre Yt: <u>Tendencia Secular: Tt:</u>

¿Por qué se debe verificar si **Y_t** observa una tendencia secular?

Wooldridge (2009: 363) comenta que no hacerlo conlleva el riesgo de que se trabaje con una **regresión espuria,** es decir que se trabaje con variables correlacionadas por casualidad más que por causalidad. Esta falsa relación de Y con t, es un problema que surge porque puede haber factores inobservables con ciertas tendencias que estén relacionados con la variable tiempo, t, los cuales, por consiguiente, influyan sin que lo sepa el investigador en la variable Y de manera que ello genera relaciones casuales (y no causales) de Y con la variable independiente, t. En otras palabras, si no se tiene cuidado en verificar lo anterior, puede suceder que por el simple hecho de observar que Y y t manifiesten tendencias de crecimiento similares puede conducir a pensar que están altamente correlacionadas, cuando en la realidad no es así.

¿Cómo identificar o aproximar la tendencia de la serie de tiempo **Y_t. ?**

Para identificar esta oscilación a largo plazo de Y_t, antes que nada se debe de trazar el **diagrama de dispersión**, mismo que puede sugerir que se corra la regresión de Y con respecto a t (tiempo). Una vez que se tiene una idea del comportamiento de Y_t se puede explorar su identificación técnica recurriendo a formas funcionales como las siguientes:

1. Lineal: $Y_t = a + bt + e_t$ se usa cuando en el diagrama de dispersión se observa que el movimiento aumenta o disminuye de manera constante, es decir, linealmente; dicha constancia la expresa el valor de la pendiente de la variable independiente y para ello se supone que $e_t = 0$, que en este caso es el tiempo, por eso en este caso se le conoce como tendencia lineal en el tiempo.

2. Cuadrática: $Y_t = a + bt + ct^2$; se sugiere utilizarlo para identificar la tendencia cuando las oscilaciones no son contantes.

3. Exponencial: Y_t = Exp $(a+b)^{t;}$ se usa cuando se vea que los valores de Y_t aumentan o disminuyen a intervalos crecientes o cuando se detecta que Y_t manifiesta el mismi ritmo (tasa) de crecimiento promedio de un lapso a otro. Cuando se decide usarla para expresar la tendencia, recuérdese que se acostumbra transformarla por medio de logaritmos para modelarla como una serie con tendencia lineal en el tiempo.

4. Logarítmica: Y_t = $\log(a+b_t)$; se utiliza cuando los valores de Y_t aumentan o disminuyen rápidamente para después terminar estabilizándose.

5. Polinomial: $Y_t = a + bt + ct^2 + \ldots + ct^n$; conviene aplicarla cuando los datos de Y_t fluctúan drásticamente. El *orden* del polinomio se determina mediante: a) el número de fluctuaciones de los valores de Y_t; ó b) de acuerdo con el número de máximos y mínimos que muestre la curva de Y_t.

6. Potencial: $Y_t = at^b$; no se usa cuando los valores de Y_t son cero o negativos; conviene aplicarla en situaciones en que se "comparan medidas que aumentan a un ritmo concreto" (Pérez, 2007: 537).

7. Tendencia de media móvil: La tendencia se identifica al trazar la curva de Y_t con al menos tres puntos: $\bar{Y}_1, \bar{Y}_2, \bar{Y}_3$. Se usa para suavizar o atenuar los valores de los términos que componen Y_t.

$$Y_t = \underbrace{\frac{Y_1 + Y_2 + Y_3}{3}}_{\bar{Y}_1} + \underbrace{\frac{Y_4 + Y_5 + Y_6}{3}}_{\bar{Y}_2} + \underbrace{\frac{X_7 + X_8 + X_9}{3}}_{\bar{Y}_3}$$

Con esta explicación didáctica de la teoría sobre el tema ahora la ilustraremos con Eviews 5, en particular <u>los métodos</u> básicos para el estudio de las series temporales con ESTACIONALIDAD

X.3.- Ejercicios con Eviews suavizamiento de las series de tiempo.

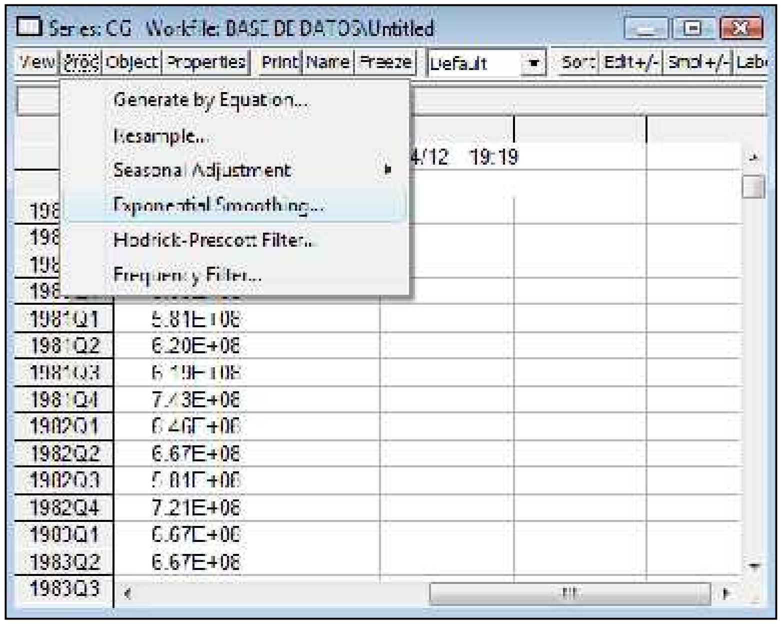
1. **Método de Brown Simple.**

¿Qué se obtiene con este método?

El método de suavización exponencial simple se usa para pronosticar una serie temporal cuando no hay tendencia o patrón estacional, pero la media de la serie temporal cambia lentamente en el tiempo.

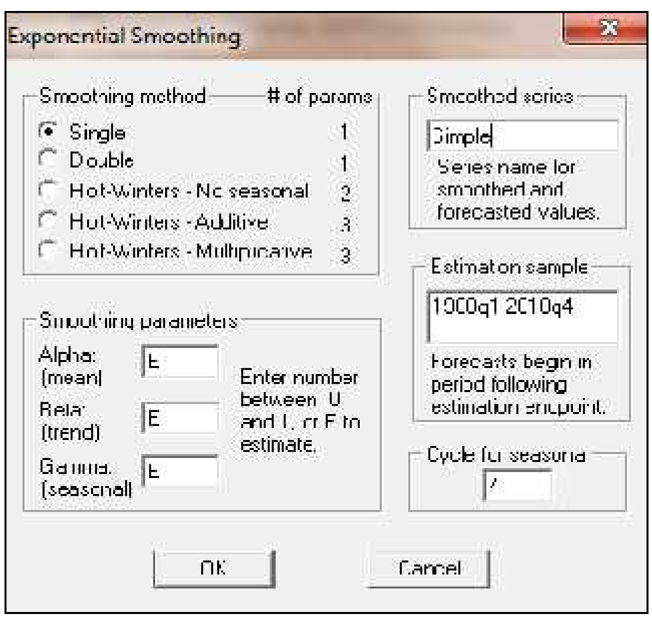
¿ Qué se busca y que relación tiene con la tendencia secular?

Utilizaremos el Consumo de Gobierno a precios constantes de 2003. Las series poseen 124 observaciones, las cuales comprenden desde el primer trimestre de 1980 hasta el cuarto trimestre de 2010. Para realizar el suavizamiento de Brown simple, seleccionamos en la serie de Consumo de Gobierno (CG) la opción Proc/ Exponential Smoothing:



Enseguida un clic, apareciendo la siguiente ventana.

Cuadro 10.3.1.1 Alisamiento Exponencial.



Se seleccionara la opción “Single” con el número de parámetro 1, dentro de los parámetros se pueden cambiar los número de alfa, beta y gamma; si no, el programa arrojará los valores. En el caso de este método de suavizamiento simple sólo se obtendrá el valor de alfa, el cual está en este rango:

$$0 \leq \alpha \leq 1$$

En caso de que “α” se encuentre más cercano a cero, significa que el nivel (media) de la serie temporal no se modifica mucho a través del tiempo. En este caso automáticamente el ciclo es 4, ya que son datos trimetrales. En la opción “smoothed series” puede escribirse el nombre que se desee que lleve la serie, en este caso escribimos “Simple”, en seguida seleccionamos “OK”. Aparecerá una ventana (veáse Cuadro 1.2.) que da los valores del parámetro alfa, la suma de errores al cuadrado , la raíz del error al cuadrado, así como el valor de la media.

Cuadro 10.3.1.2. Análisis del método exponencial Brown simple.

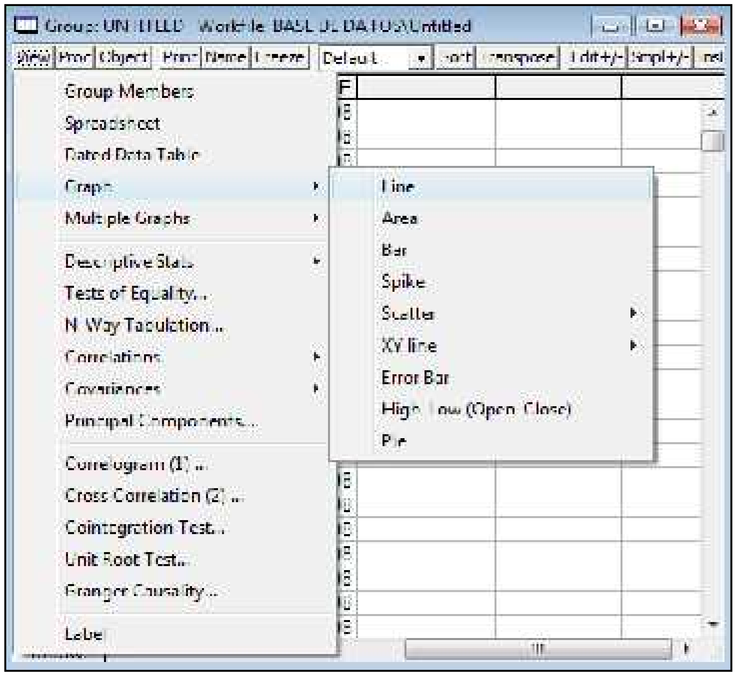
Sample: 1980Q1 2010Q4
Included observations: 124
Method: Single Exponential
Original Series: CG
Forecast Series: SIMPLE

Parameters:	Alpha	0.1060
Sum of Squared Residuals		1.65E+18
Root Mean Squared Error		1.15E+08
End of Period Levels:	Mean	9.73E+08

¿Cuál es la interpretación de estos resultados?

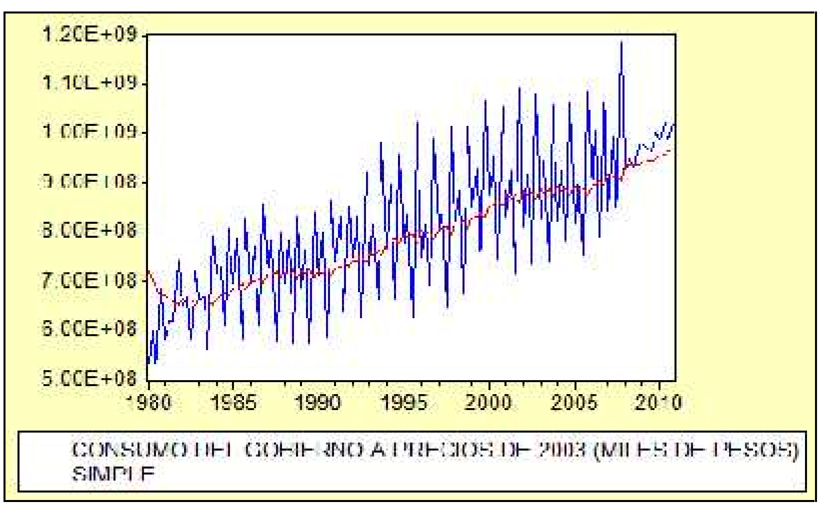
El parámetro “Alpha” no se encuentra cerca al cero, por lo cual se considera que persisten cambios en el tiempo. Lo cual puede observarse en el valor de la media final que es de 967, 520,117.40 y para iniciar el pronóstico fue de 721, 451,489.08, puede apreciarse que no ha permanecido la media constante. Por lo cual puede que este método de suavizamiento no sea el correcto para realizar un pronóstico.

Para observar el comportamiento de la serie de Consumo de Gobierno y Consumo de Gobierno con el método de Brown simple, en la pantalla de Workfile se seleccionan ambas variables “cg” y “simple”, se da un clic derecho, se selecciona Open/ as group…, apareciendo una ventana con los valores de las dos variables. En seguida se selecciona View/ Graph/ Line



Apareciendo la siguiente gráfica.

Cuadro 10.3.1.3 Gráfica de Consumo de Gobierno y Consumo de Gobierno Simple.

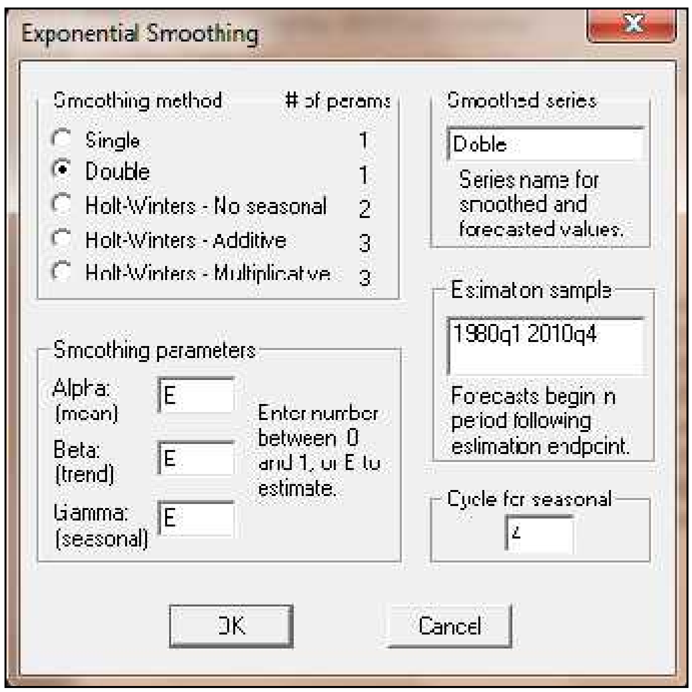


De esta manera podemos observar que el pronóstico que realiza esta prueba no se ajusta a la serie de tiempo del Consumo de Gobierno, lo cual ya habíamos analizado en el parámetro alfa y el comportamiento de la media inicial con la media final, además de que este método supone una serie temporal sin tendencial lo cual no sucede en el comportamiento de esta variable.

2. **Método de Brown Doble.**

Derivado de lo anterior es que en seguida se utilizará el método doble de Brown, con este método de ensanchamiento se reducen las variaciones y por ende se logra un mejor alisamiento de la curva. Para la utilización de este método en la barra de comandos se escribe “show cg”, apareciendo la serie de Consumo de Gobierno, seleccionamos dentro de la barra de herramientas Proc/ Exponetial Smoothing y enseguida un clic, apareciendo la siguiente ventana.

Cuadro 10.3.2.1. Alisamiento Exponencial.



Para este ejemplo, se seleccionará la opción “Double” con el número de parámetro 1, dentro de los parámetros alisados se deja “E” y el valor lo dará el programa. Se mantiene el mismo ciclo. En la opción “smoothed series” escribimos “Doble”. En seguida sale una ventana que da los valores del parámetro alfa, la suma de errores al cuadrado, la raíz del error al cuadrado, así como el valor de la media y la tendencia.

Cuadro 10.3.2.2 Análisis del método exponencial Brown doble.

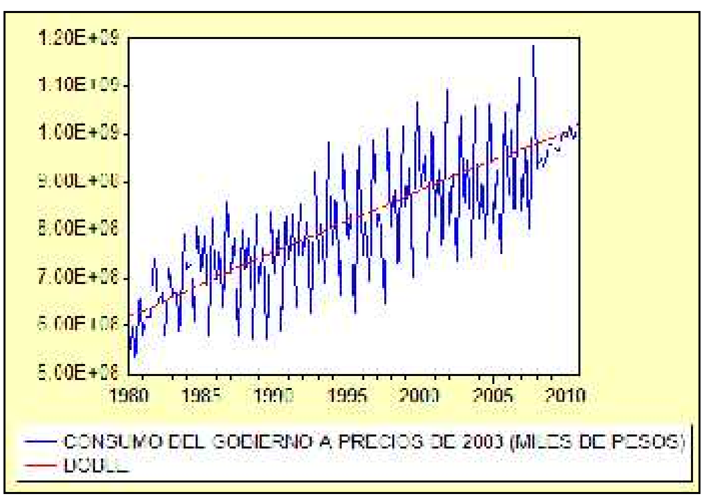
Sample: 1980Q1 2010Q4
Included observations: 124
Method: Double Exponential
Original Series: CG
Forecast Series: DOBLE

Parameters: Alpha		0.0050
Sum of Squared Residuals		1.41E+18
Root Mean Squared Error		1.07E+08
End of Period Levels:	Mean	1.01E+09
	Trend	3282612.

El parámetro “Alpha” se encuentra cerca al cero, por lo cual se considera que no persisten cambios en el tiempo. Lo cual puede observarse en el valor de la media final que es de 1, 014, 891,579.01 y para iniciar el pronóstico fue de 620, 411,289.07, los resultados son diferentes del método anterior, debido a que se está considerando el componente de tendencia en la serie de tiempo del Consumo de Gobierno. Por lo cual puede que este método de suavizamiento doble de Brown sigue sin ser el más adecuado, aunque para está serie es más indicado que el método simple de Brown.

Para observar el comportamiento de la serie de Consumo de Gobierno y Consumo de Gobierno con el método de Brown doble, en la pantalla de Workfile se seccionan ambas variables “cg” y “doble”, se da un clic derecho, se selecciona Open/ as group…, apareciendo una ventana con los valores de las dos variables. En seguida se selecciona View/ Graph/ Line, apareciendo la siguiente gráfica.

Cuadro 10.3.2.3 Gráfica de Consumo de Gobierno y Consumo de Gobierno Doble.

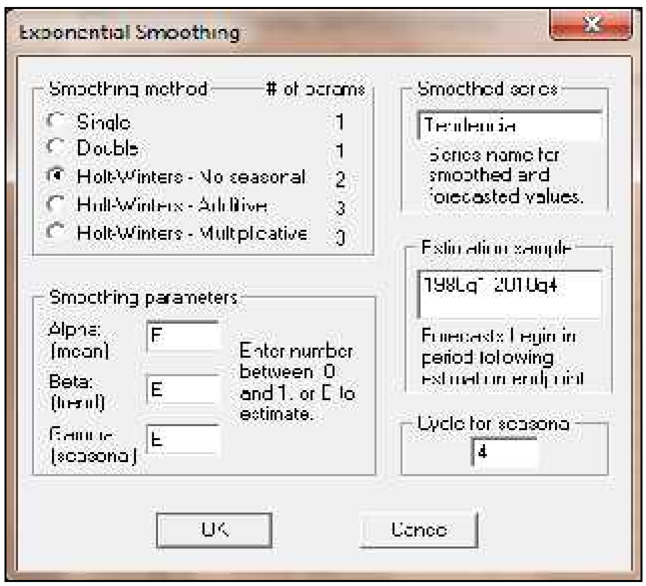


De igual forma que el método del inicio podemos observar que el pronóstico que realiza esta prueba no se ajusta a la serie de tiempo del Consumo de Gobierno, aunque este método supone una serie temporal con tendencia.

3. **Método de Holt-Winters.**

Si la serie temporal manifiesta una tendencia lineal, en caso de que incremente o disminuya su tasa aproximadamente constante, entonces el modelo de tendencial lineal es el método más apropiado cuando cambia la media y la tasa de crecimiento. Para este método se siguien los mismo pasos para la obtención de la ventana de "Alisado Exponencial". En la barra de comandos se escribe "show cg", seleccionamos dentro de la barra de herramientas Proc/ Exponential Smoothing y enseguida un clic, aparece lo siguiente.

Cuadro 10.3.3.1 Alisamiento Exponencial.



Se seleccionará la opción "Holt-Winters - No seasonal" con el número de parámetro 2, dentro de los parámetros alisados se deja "E" y el valor lo dará automáticamente el programa. Se mantiene el mismo ciclo. En la opción "smoothed series" escribimos "Tendencia". Enseguida sale una ventana que da los valores de los parámetros alfa y beta, la suma de errores al cuadrado, la raíz del error al cuadrado, así como el valor de la media y la tendencia.

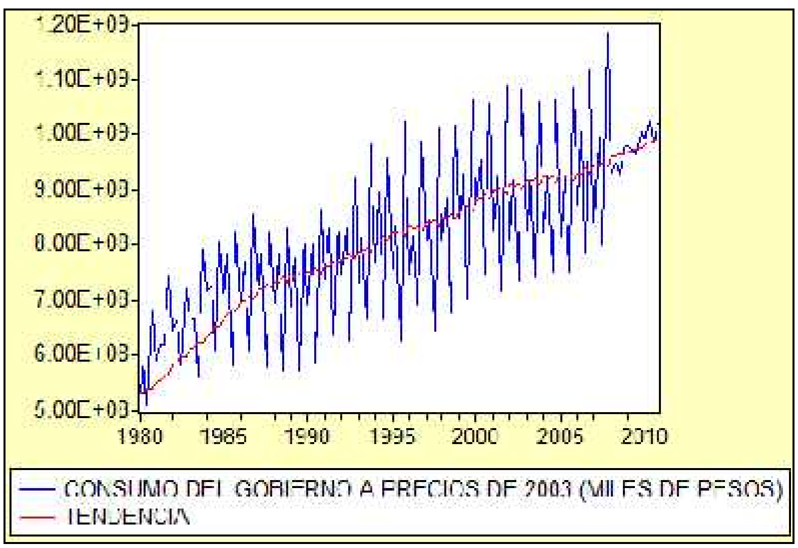
Cuadro 10.3.3.2 Análisis del método exponencial Holt-Winters – No seasonal.

Sample: 1980Q1 2010Q4
Included observations: 124
Method: Holt-Winters No Seasonal
Original Series: CG
Forecast Series: TENDENCIA

Parameters:	Alpha	0.0700
	Beta	0.0200
Sum of Squared Residuals		1.55E+18
Root Mean Squared Error		1.12E+08
End of Period Levels:	Mean	9.96E+08
	Trend	3078688.

Para observar el comportamiento de la serie de Consumo de Gobierno y Consumo de Gobierno con el método de Holt-Winters, en la pantalla de Workfile se seleccionan ambas variables “cg” y “tendencia”, se da un clic derecho, se selecciona Open/ as group…, apareciendo una ventana con los valores de las dos variables. Enseguida se selecciona View/ Graph/ Line, apareciendo la siguiente gráfica.

Cuadro 10.3.3.3 Gráfica de Consumo de Gobierno y Consumo de Gobierno Tendencia.



Los valores iniciales no se ajustan muy bien a los datos lo que arroja una suma de errores al cuadrado grande (Véase cuadro 3.2). Como resultado, los intervalos de predicción son más amplios de lo necesario. Asimismo, por este método podemos observar que el pronóstico que realiza esta prueba no se ajusta a la serie de tiempo del Consumo de Gobierno, aunque este método supone al igual que método Doble de

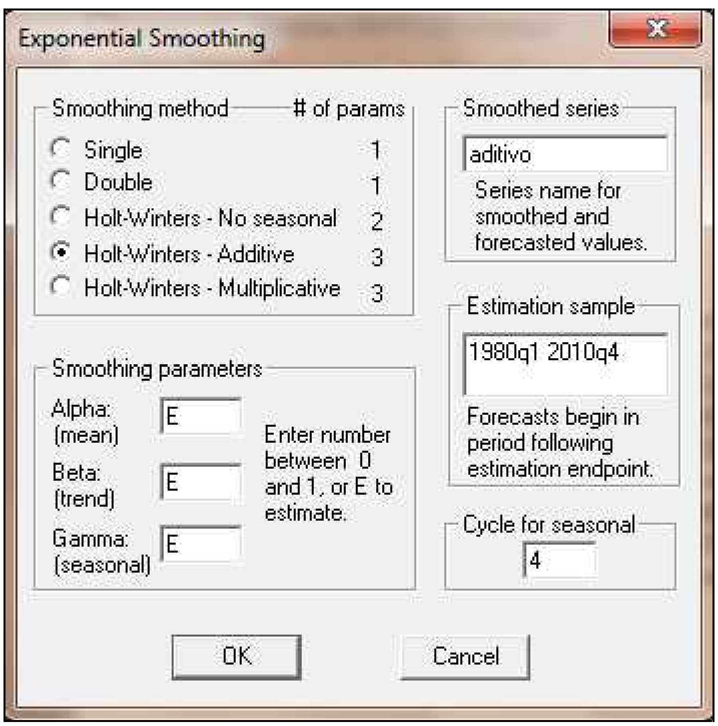
Brown una serie temporal con tendencia y su ajuste es más adecuado, pero no el mejor.

Método de Holt-Winters Aditivo.

El método de Holt-Winters aditivo se usa para las series de tiempo que presentan una variación estacional fija, además muestran una tendencial lineal con una tasa de crecimiento constante. Es adecuado cuando la serie temporal muestra una tendencial lineal con un patrón estacional aditivo, en donde el nivel, la tasa de crecimiento y el factor estacional podrían estar cambiando.

Para este método se siguien los mismo pasos para la obtención de la ventana de "Alisado Exponencial". En la barra de comandos se escribe "show cg", seleccionamos dentro de la barra de herramientas Proc/ Exponetial Smoothing y enseguida un clic, aparece lo siguiente.

Cuadro 10.3.3.4 Alisamiento Exponencial.



Se seleccionará la opción "Holt-Winters - Additive" con el número de parámetro 3, dentro de los parámetros alisados (α, γ y δ son constantes de suavización sus valores se encuentran entre 0 y 1) se deja "E" y los valores los dará automáticamente el programa, se mantiene el mismo ciclo. En la opción "smoothed series" escribimos "Aditivo". En seguida en la ventana de análisis del método exponencial Holt-Winters aditivo (véase Cuadro 4.2) que da los valores de los parámetros alfa, beta y gamma, la suma de errores al cuadrado, la raíz del error al cuadrado, así como el valor de la media, la tendencia y el ciclo.

Cuadro 10.3.3.5 Análisis del método exponencial Holt-Winters Aditivo

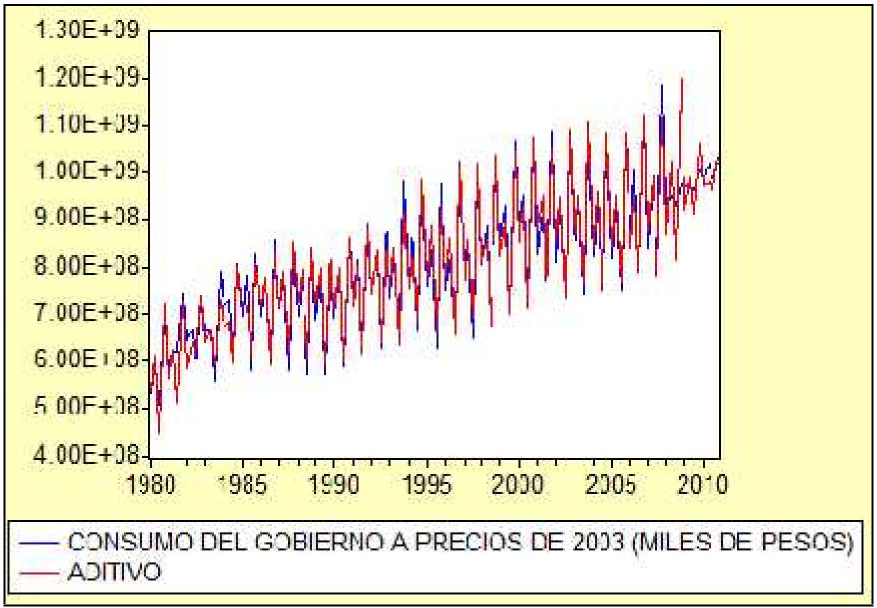
Sample: 1980Q1 2010Q4
Included observations: 124
Method: Holt-Winters Additive Seasonal
Original Series: CG
Forecast Series: ADITIVO

Parameters:	Alpha			0.1700
	Beta			0.0000
	Gamma			0.8102
Sum of Squared Residuals				1.97E+17
Root Mean Squared Error				39840280
End of Period Levels:	Mean			1.00E+09
	Trend			3541222
	Seasonals:	2010Q1		-4559491
		2010Q2		14242055
		2010Q3		-24023623
		2010Q4		14341058

A diferencia de los métodos pasados en el de Holt-Winters aditivo se considera la tendencia y el componente estacional. Lo que puede concluirse es que las constates de suavización (α, γ, δ) para la tasa de crecimiento son cero, por lo tanto, las estimaciones no son distintas de las estimaciones iniciales.

Para observar el comportamiento de la serie de Consumo de Gobierno y Consumo de Gobierno con el método de Holt-Winters aditivo, en la pantalla de Workfile se seccionan ambas variables “cg” y “aditivo”, se da un clic derecho, se selecciona Open/ as group…, apareciendo una ventana con los valores de las dos variables. En seguida se selecciona View/ Graph/ Line, apareciendo la siguiente gráfica.

Cuadro 10.3.3.6 Gráfica de Consumo de Gobierno y Consumo de Gobierno Aditivo.



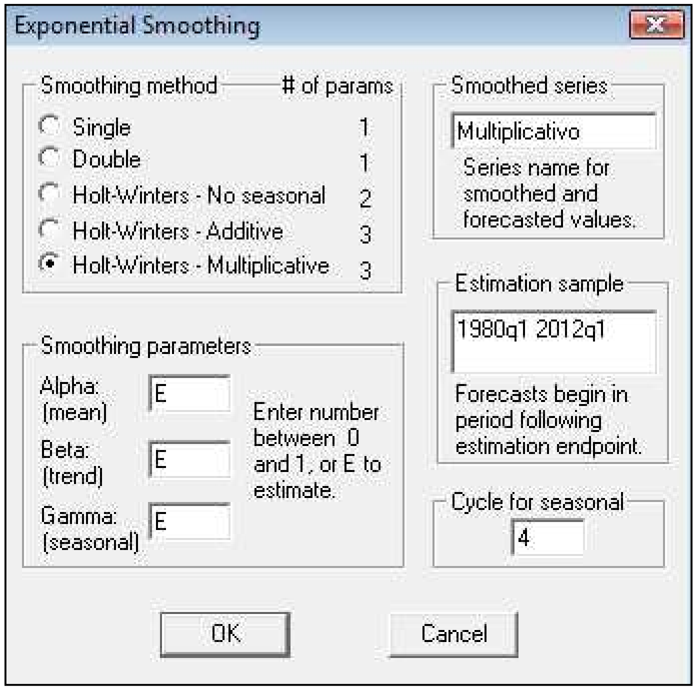
Interpretación: Con la gráfica es posible ver como el método de suavizamiento exponencial de Holt-Winters aditivo se ajusta de mejor forma que en los métodos anteriores, esto porque considera más parámetros, así como el componente estacional y la tendencia. No puede determinarse aún si es el mejor método de suavizamiento debido a que se aplica a una serie temporal fija.

4. **Método de Holt-Winters Multiplicativo.**

En este método la serie temporal muestra una tendencia lineal, así como una tasa de crecimiento lineal y un patrón estacional constante. Es adecuada cuando una serie muestra una tendencia lineal y un factor estacional multiplicativo, para la cual puede que el nivel, la tasa de crecimiento y la variación estacionalidad podrían ser fluctuantes.

Para este método se siguien los mismo pasos para la obtención de la ventana de “Alisado Exponencial”. En la barra de comandos se escribe “show cg”, seleccionamos dentro de la barra de herramientas Proc/ Exponetial Smoothing y enseguida un clic, aparece lo siguiente.

Cuadro 10.3.4.1 Alisamiento Exponencial.



Se seleccionará la opción “Holt-Winters - Multiplicative” con el número de parámetro 3, dentro de los parámetros alisados se deja “E” y los valores los dará automáticamente el programa, se mantiene el mismo ciclo. En la opción “smoothed series” escribimos “Multiplicativo”. Enseguida sale una ventana análisis del método exponencial Holt-Winters Multiplicativo (véase cuadro 5.2) que da los valores de los parámetros alfa, beta y gamma, la suma de errores al cuadrado, la raíz del error al cuadrado, así como el valor de la media, la tendencia y el ciclo.

Cuadro 10.3.4.2 Análisis del método exponencial Holt-Winters Multiplicativo.

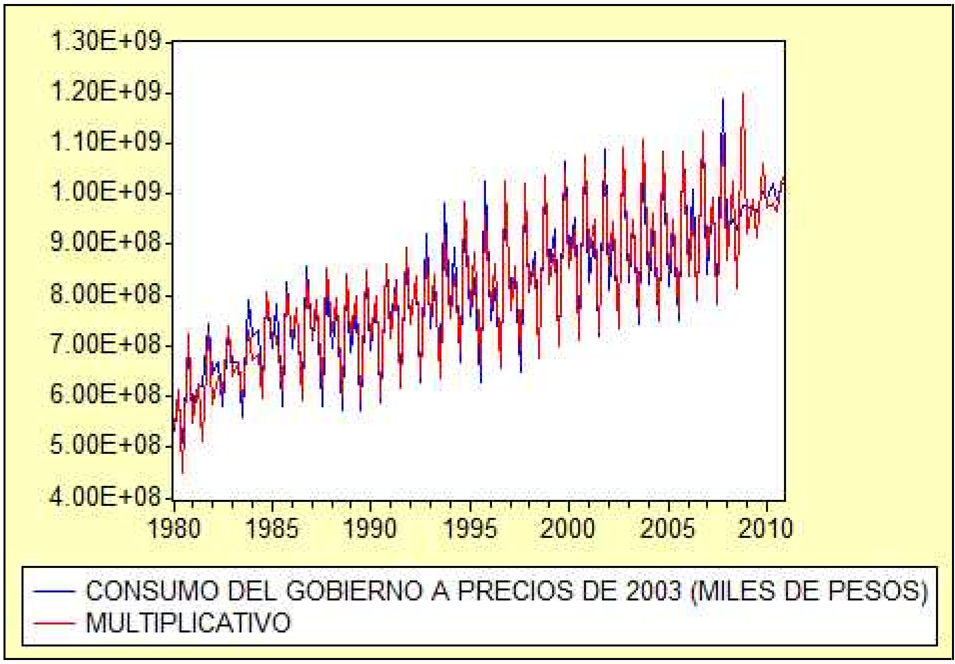
Sample: 1980Q1 2010Q4
Included observations: 124
Method: Holt-Winters Multiplicative Seasonal
Original Series: CG
Forecast Series: MULTIPLICATIVO

Parameters:	Alpha		0.1600
	Beta		0.0000
	Gamma		0.7702
Sum of Squared Residuals			1.97E+17
Root Mean Squared Error			39883141
End of Period Levels:	Mean		1.00E+09
	Trend		3541222.
	Seasonals:	2010Q1	0.993593
		2010Q2	1.014003
		2010Q3	0.975318
		2010Q4	1.017086

Los resultado que arroja este método exponencial Holt-Winters multiplicativo son similares a los del métodos expencial de Holt-Winters aditivo. Lo que puede concluirse es que las constates de suavización (α, γ, δ) para la tasa de crecimiento son cero, por lo tanto, las estimaciones no son distintas de las estimaciones iniciales. Pero existe una suma de errores al cuadrado grande lo que refleja la lorgitud de los intervalos en la predicción, los intervalos son mucho más amplios.

Para observar el comportamiento de la serie de Consumo de Gobierno y Consumo de Gobierno con el método de Holt-Winters multiplicativo, en la pantalla de Workfile se seccionan ambas variables “cg” y “multiplicativo”, se da un clic derecho, se selecciona Open/ as group…, apareciendo una ventana con los valores de las dos variables. En seguida se selecciona View/ Graph/ Line, apareciendo la siguiente gráfica.

Cuadro 10.3.4.1 Gráfica de Consumo de Gobierno y Consumo de Gobierno Multiplicativo.



¿Qué se logró con este ejercicio?

Con los datos generados para el suavizamiento exponencial Holt-Winters multiplicativo se obtinen una gráfica se ajusta de mejor forma que en los métodos anteriores, esto porque al igual que el método anterior considera más parámetros, así como el componente estacional y la tendencia. Aunque podría establecer que la serie de Consumo de Gobierno muestra una tendencia lineal con una tasa de crecimiento lineal y un patrón estacional constate que puede cambiar con el paso del tiempo.

X.4.-PROFUNDIZACIÓN SOBRE LOS COMPONENTES Y ANÁLISIS DE SERIES DE TIEMPO

Una serie de tiempo es una colección de observaciones de una misma variable realizadas de forma secuencial a través del tiempo, pueden ser considerados en meses, trimestres, años, etc.

Matemáticamente, las observaciones se denotan como:

$$Xt = (x_{1,}\ x_2, x_3 \ldots x_k)$$

Una serie de tiempo puede:

$(x_t \mid x_k \in Z)$ Entonces es una serie de tiempo discreta.

$(x_t \mid x_k \in R)$ En este caso se trata de una serie de tiempo continua.

Los posibles objetivos del análisis de series de tiempo, de acuerdo con Chatfield son: descripción, explicación, predicción y control.

Descripción: Dotará el análisis con los componentes principales de la serie. Las cuales se presentan a continuación:

1. **Ciclicidad**. Este concepto se refiere a variaciones de la serie en periodos mayores a un año. Altas y bajas de las observaciones que suelen repetirse. El componente cíclico es la fluctuación en forma de onda alrededor de la tendencia. **La ciclicidad es un fenómeno que en lo general parece estar relacionado con la variación de la actividad económica ocurrida durante periodos de crisis o prosperidad. La fluctuación también puede presentarse en series de tiempo estacionarias.**
2. **Estacionalidad**. Se refiere a las variaciones que tiene la serie de tiempo en un periodo menor a un año y que tienden a repetirse a sí mismas año con año. **El patrón de cambio por lo general es un aumento o una disminución cuantitativa en los valores observados de una serie de tiempo específica. Cabe mencionar que aunque en la mayor parte de los casos el patrón estacional es un fenómeno que se presenta en lapsos de tiempo de duración aproximada a un año; también puede manifestarse éste fenómeno en periodos de tiempo, ya sean menores o mayores a un año.**
3. **Tendencia**. Esta fluctuación se puede identificar a largo plazo e indica el crecimiento o decrecimiento de la serie.

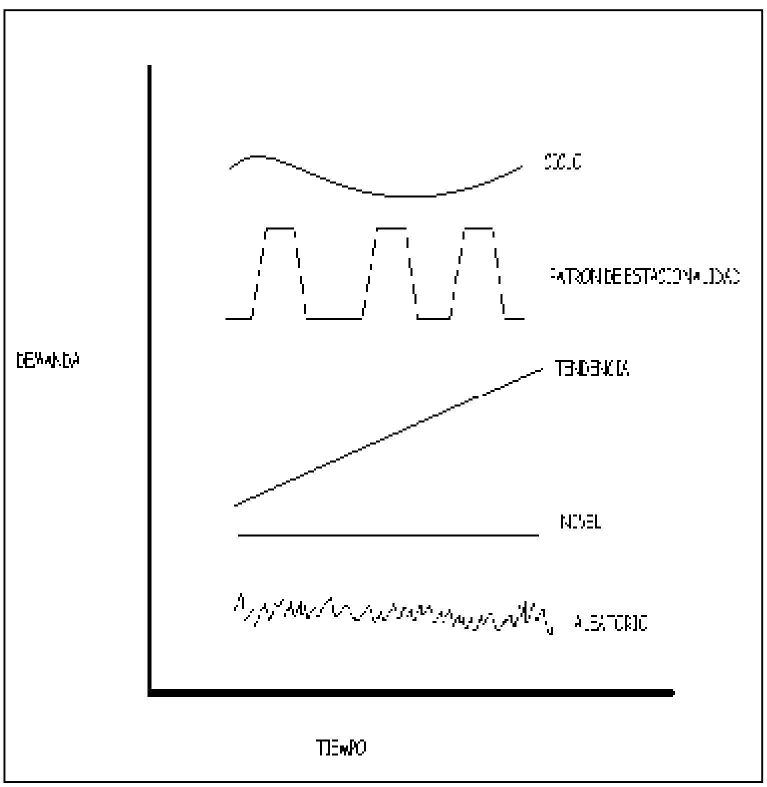
   La tendencia de una serie de tiempo es el componente de largo plazo que representa el crecimiento o disminución en la serie sobre un periodo amplio. La tendencia es la propensión al aumento o disminución en los valores de los datos de una serie de tiempo, que permanece a lo largo de

un lapso muy extendido de tiempo, es decir que no cambiará en el futuro lejano mientras no hayan cambios significativos o radicales en el entorno en el que se encuentra inmersa y que determina el comportamiento de la serie de tiempo en estudio, cambios que podrían ser originados como por ejemplo, por descubrimientos científicos, avances tecnológicos, cambios culturales, geopolíticos, demográficos, religiosos, etc.**

4. **Aleatoriedad.** Al eliminar los otros componentes la variabilidad restante de las series de tiempo suele ser la de índole aleatoria. Esta puede ser provocado por sucesos externos e imprevisibles. **La aleatoriedad es el cambio producido en los valores de una serie de tiempo debido a fenómenos que son en extremo difíciles de explicar y que por lo tanto su ocurrencia cae en el ámbito del azar.**

Gráficamente las fluctuaciones de la serie se muestran en el cuadro siguiente.

Gráfica 10.4.1. Representación de las fluctuaciones de una serie de tiempo.



X.4.1. Identificación con datos mexicanos de Variaciones Cíclicas, Ct arriba descritas

Es importante reiterar que este tipo de oscilación o movimiento de la serie temporal es mucho más difícil de identificar dado que a diferencia de la tendencia, T_t, que es una variación constante observada en el largo plazo, o de una variación estacional; E_t, que tiene un lapso fijo, las oscilaciones cíclicas; C_t, no tienen un periodo fácilmente identificable, que en *ciertas ocasiones* inclusive es *variable*; este último caso origina que con frecuencia coexistan ciclos que se superponen entre si. Esta situación hace aun más difícil identificar las variaciones cíclicas de una serie temporal, Y_t.

Como antes se dijo, los ciclos generalmente tienen cuatro movimientos o etapas, si se inician por ejemplo, con la de recuperación, le sigue la de prosperidad o auge, a la que le sucede la recesión y luego la depresión, en periodos mayores a un año. Los movimientos de estas etapas fluctúan por abajo y por arriba de la tendencia secular (ver gráfica 10.4.1).

Derivado de lo anterior es que para detectar el ciclo, en forma práctica, se acostumbra, de manera secuenciada, primero eliminar de la serie original la tendencia y, enseguida, las oscilaciones estacionales, de manera que después se analiza el *resto* (χ) de la serie, tal que suele expresarse así:

$$X_{ik} = C_{ik} + r_{ik}$$

En ocasiones se prescinde del doble subíndice ya que en este tipo de variaciones no existen los de carácter estacional.

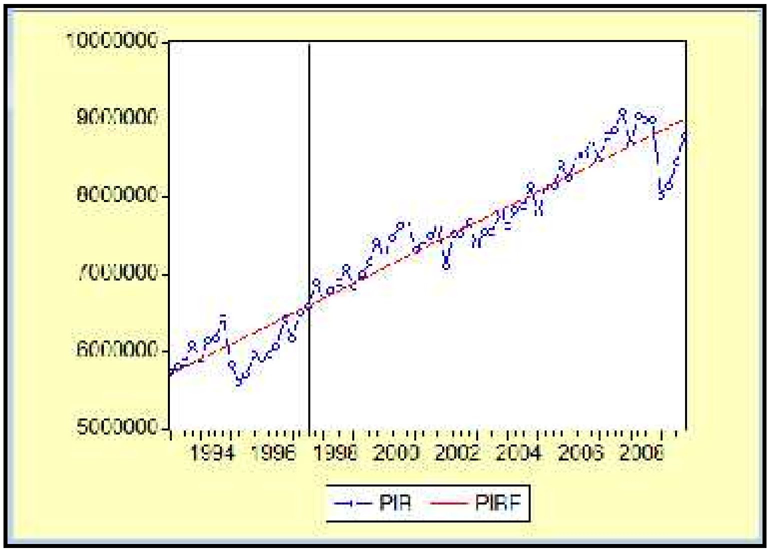
El algoritmo que se usa para identificar las variaciones cíclicas en la nueva serie χ, se conoce como el *método de análisis armónico*, que luego se ilustrará, del cual también puede derivarse otro método: el del *periodograma acumulativo*.

Ejemplo:

De la serie del PIB anteriormente utilizada para identificar el componente de tendencia, ahora también puede observarse claramente que ésta presenta un *claro comportamiento cíclico,* en la cual pueden observarse las cuatro etapas características de un ciclo, a saber: 1.- *recuperación; 2.- auge; 3.- recesión y 4.- crisis.*

Así, en la gráfica siguiente 1.2 pueden identificarse claramente **tres ciclos**, cada uno de ellos con sus respectivas 4 etapas. El primer ciclo parece iniciar un poco antes del primer trimestre de 1993 y termina en el tercer trimestre de 1997. El segundo ciclo inicia en el tercer trimestre de 1997 y termina en el segundo o tercer trimestre de 2005. El tercer ciclo inicia en el tercer trimestre de 2005 y a la fecha aún no ha concluido.

Gráfica 10.4.1.1: Ciclos observados



10.4.2. Identificación de Variaciones Estacionales, Et.

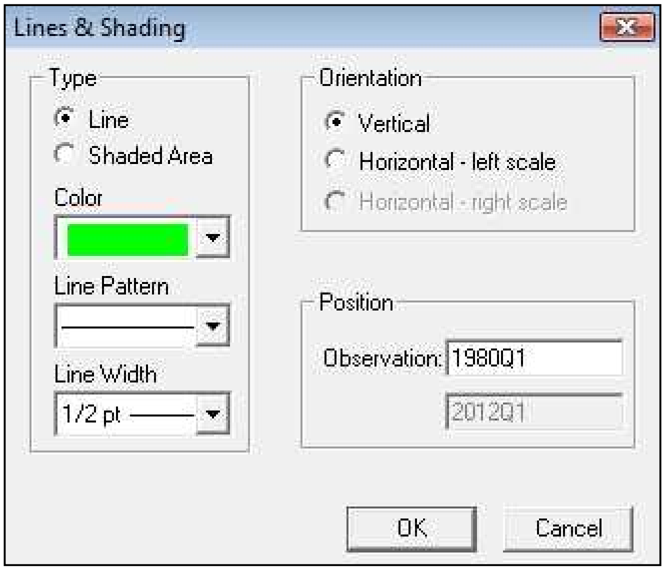
Se *identifican* graficando los datos con algunas de las formas funcionales antes descritas y detectando que *algunos de ellos se reproducen en cada uno* de los años de Y_t.

Ejemplo:

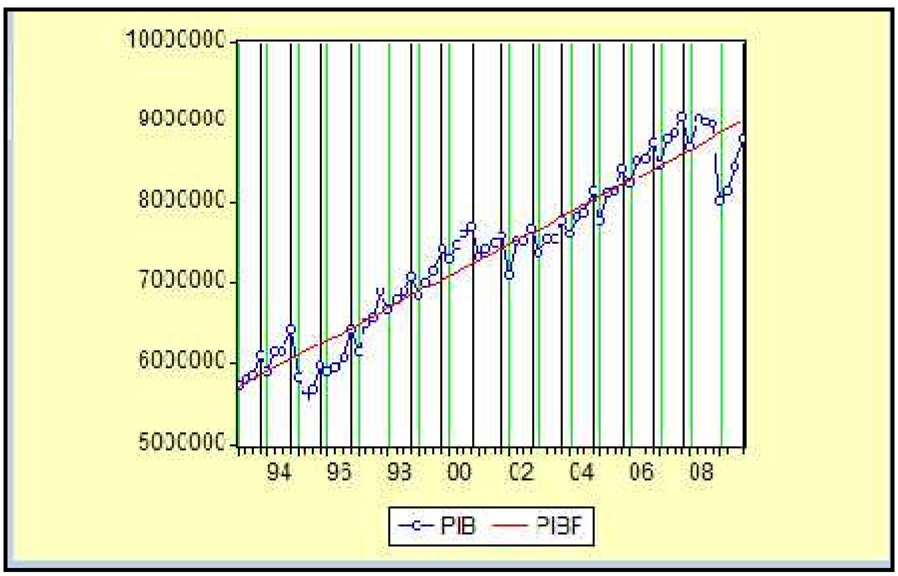
De la serie del PIB puede observarse que cada primer trimestre de cada año el PIB disminuye y durante el cuarto trimestre de cada año el PIB aumenta, ello es un claro ejemplo de estacionalidad. Sin embargo, no todas las series, con frecuencia mensual o trimestral, presentan estacionalidad.

Así, en la Gráfica 1.3 siguiente las líneas verticales de color verde muestran la disminución del PIB durante cada primer trimestre de cada año y las líneas verticales de color negro muestran el aumento del PIB durante cada cuarto trimestre de cada año.

Por otra parte, para agregar las líneas verticales en Eviews una vez que se ha generado la gráfica sobre ésta se da un click derecho y se selecciona la opción de “add shading” y en la ventana que se desplegará en “orientation” seleccionamos vertical y en “position” tenemos que ir escribiendo cada una de las observaciones que presentan el comportamiento estacional, en nuestro caso: 1993Q1, 1994Q1,…, y 2009Q1 para el primer trimestre de cada año y 1993Q4, 1994Q4,…, y 2009Q4 para el cuarto trimestre de cada año. Nota: Deben ponerse una por una cada una de las observaciones, también puede elegirse el color de la línea, la forma de la línea y el grosor de la misma.



Gráfica 10.4.1.2: Identificación de variaciones estacionales



X.4.2.1 ¿Cómo se eliminan las variaciones estacionales?

En virtud de que la <u>estacionalidad</u> observada en los datos periódicamente produce un efecto distorsionador del movimiento armónico de Y_t, es necesario eliminarla para conocer la oscilación real de Y_t. y estar en condiciones de enseguida predecir los valores futuros de la serie temporal con mayor certeza.

Ello se logra con la “desestacionalización” de la serie, que se puede obtener con *métodos* sencillos como: a) método de la tendencia; b) de las medias móviles; y c) el método de las diferencias estacionales. Estos métodos se describen a continuación:

X.4.2.1.1.-Método de la tendencia: también conocido como método de las relaciones de medias mensuales respecto a la tendencia.

Notación: Si la serie temporal se representa con Y_{ik} y si establecemos que i representa el año i-ésimo tal que i = 1, 2, 3,…N y si definimos a k como la estación késima del año i-ésimo, donde k = 1, 2, 3,…m. Observe que: a) cuando la estación del año está compuesto por meses, entonces m = 12 y b) cuando el año está compuesto por trimestres, entonces m = 4. Invariablemente T = Nm, cuando suponemos que los términos de la serie están ordenados uno tras otro según van apareciendo en el tiempo, de manera que t = 1, 2, 3,…, T por lo que T = Nm.

Pasos para eliminar o desestacionalizar la tendencia estacional que observa la serie histórica en el tiempo.

1. Calcular las *medias anuales* de los datos y con ellas hacer la regresión:

$$\bar{Y}_i = a - b_i$$

Donde:

$$\bar{Y}_i = \frac{1}{m}\sum_{k=1}^{m} Y_{ik}$$

2. Obtener las *medias mensuales* en cada uno de los años:

$$\bar{Y}_k = \frac{1}{N}\sum_{i=1}^{N} Y_{ik}$$

Donde: k = 1, 2, 3,…, m.

3. Aislar el componente estacional obteniendo la serie de medias mensuales *corregidas*:

$$\overline{Y}_k' = \overline{Y}_k - \frac{b(k-1)}{m}$$

4. Calcular la media global corregida:

$$\overline{Y}' = \frac{\overline{Y}_1' + \overline{Y}_2' + \ldots + \overline{Y}_m'}{m}$$

5. Si el esquema o proceso es *multiplicativo*, entonces se calculan los índices de variación estacional:

$$I_K = \frac{\overline{Y}_k'}{\overline{Y}'}100$$

   Así, si dividimos los valores de la serie entre las I_k, ésta queda desestacionalizada, sabiendo que la componente estacional es:

$$E_{ik} = \frac{I_k}{100} = \frac{\overline{Y}_k'}{\overline{Y}'}$$

6. Cuando el proceso o esquema es aditivo, el componente del mes k es:

$$E_{ik} = \overline{Y}_k' - \overline{Y}'$$

X.4.2.1.2.- Método de desestacionalización de las medias móviles.

En opinión de César Pérez (2007: 540) este método "consiste en obtener la componente extraestacional mediante un ajuste de la serie original por medias móviles de orden m para eliminar las variaciones estacionales".

Ejemplo:

Suponga nuevamente la serie del PIB de México para el periodo comprendido entre el primer trimestre de 1993 y el cuarto de 2009 a precios constantes de 2003. Como ya se determino en el punto XI. 13. 3, esta serie presenta estacionalidad pues decrece en todos los primeros trimestres y repunta en todos los cuartos trimestres, por lo que este comportamiento se suavizará utilizando el método del promedio o media móvil. Este método consiste en “mover los valores de la media aritmética a través de la serie del tiempo” (Mason, et al.-, 2000: 655).

Pasos para emplear el método de medias móviles:

1. Determinar el valor de m, donde éste se refiere al número de observaciones que hay en un año, es decir, la frecuencia de los datos. Así, si la frecuencia es mensual, m = 12; si es trimestral, m = 4; etc.

2. Calcular la media aritmética de los “m” primeros valores de la serie y colocar el resultado a la altura de la observación central. Aquí se presentan dos casos:

   a) La frecuencia de los datos es impar: En este caso la media móvil se colocará en la observación central. Por ejemplo: si la frecuencia de los datos es m = 3, entonces la media se colocará en la observación 2; si m = 5, la media se colocará en la observación 3; etc.

   b) La frecuencia de los datos es par: Como en el caso anterior la media móvil se colocará en la observación central, sin embargo, como m es par la media móvil se colocará entre las dos observaciones centrales, teniendo además que calcular la media móvil centrada. Por ejemplo: si m = 4, entonces la media móvil se colocará entre la observación 2 y 3, hasta completar toda la serie, después se calculará la media móvil centrada de la serie precedente, quedando ahora la media en la tercera observación de cada año. Este caso se ejemplificará a continuación.

3. Repetir el paso 2 eliminando el primer valor de la serie, por ejemplo, supongamos que m = 3 y que tenemos los años 2000 y 2001, entonces tendríamos valores para las observaciones 1, 2, 3, 4, 5 y 6. La primer media móvil será:

$$\ddot{x} = \frac{valor\ \ 1 + valor\ \ 2 + valor\ \ 3}{3}$$

   La segunda media móvil será:

$$\ddot{x} = \frac{valor\ \ 2 + valor\ \ 3 + valor\ \ 4}{3}$$

   La tercera media móvil será:

$$\bar{x} = \frac{valor \quad 3 + valor \quad 4 + valor \quad 5}{3}$$

Y así sucesivamente.

4. La serie que se obtenga será la serie desestacionalizada. Sin embargo, si m es impar se perderán m-1 observaciones, por ejemplo, si m = 3 se perderán dos observaciones; si m = 5, se perderán 4 observaciones, etc.,; si m es par se perderán m observaciones, por ejemplo, si m = 4 se perderán 4 observaciones; si m = 12 se perderán 12 observaciones, etc.

En nuestro caso como vamos a trabajar con la serie del PIB de forma trimestral, m es par, por lo que tendremos que calcular la media móvil de la serie y además, a partir de ésta, se calculará la medía móvil centrada. Así, con fines ilustrativos y por cuestión de espacio únicamente se muestra una parte de las operaciones aritméticas necesarias para obtener la serie desestacionalizada. Los cálculos se realizaron en Excel insertando filas donde fuera oportuno y/o necesario.

Así, observe el siguiente cuadro y la explicación siguiente:

Imagen 10.4.2.1.2.1 Obtencion de datos

	A	B	C	D	E	F
16						
17		Periodo	PIB	Media móvil	Media Móvil Centrada	
18		1993/01	5732604.63			
19		1993/02	5802801.77			
20				5871531.02		
21		1993/03	5657030.53		5891950.99	
22				5912390.96		
23		1993/04	6092879.1		5954406.99	
24				5996423.03		
25		1994/01	5696044.37		6033328.06	
26				6070233.1		
27		1994/02	6138930.06		6111726.19	
28				6153219.27		
29		1994/03	6153078.88		6144433.21	
30				6135647.15		
31		1994/04	6424823.79		6068276.25	
32				6000905.34		
33		1995/01	5825755.00		5943259.7	

Como se comentó anteriormente, en este caso m = 4, por tanto la frecuencia es par y se tendrá que calcular la media móvil y la media móvil centrada.

a) Cálculos para la serie de la media móvil:

Como m es par e igual a 4, la media móvil se colocará entre las observaciones 2 y 3, en este caso entre el segundo y el tercer trimestre de 1993. Así, la primera media móvil será:

$$\bar{x}_1 = \frac{5732604.63 + 5802801.76 + 5857838.58 + 6092879.10}{4} = 5871531.02$$

Cuyo valor se colocará entre las observaciones 1993/02 y 1993/03 (obsérvese fila 20 columna D del cuadro anterior).

Para la segunda media móvil omitimos el primer valor de la serie y tomamos los siguientes cuatro, es decir, en este caso, los valores de las observaciones 1993/02, 1993/03, 1993/04 y 1994/01, obteniéndose la siguiente media móvil:

$$\bar{x}_2 = \frac{5802801.76 + 5857838.58 + 6092879.10 + 5896044.37}{4} = 5912390.96$$

Cuyo valor se colocará entre las observaciones 1993/03 y 1993/04 (observe el cuadro anterior).

Para la tercera media móvil omitimos ahora tanto el primero como el segundo valor de la serie y tomamos los siguientes cuatro, es decir, los valores de las observaciones 1993/03, 1993/04, 1994/01 y 1994/02, obteniendo:

$$\bar{x}_3 = \frac{5857838.58 + 6092879.10 + 5896044.37 + 6138930.04}{4} = 5996423.03$$

Y así sucesivamente hasta completar la serie, en este caso, la de medias móviles, si m fuera impar únicamente tendría que calcularse está serie, pero dado que m es par entonces se tendrá que calcular la serie de la media móvil centrada.

b) Cálculos para la serie de la media móvil centrada:

Una vez que contemos con la serie de la media móvil, *ahora centraremos su valo*r. Así, se procede a calcular y centrar la media de dos medias móviles. En nuestro caso, calcularemos la media de las medias móviles que se encuentran entre las observaciones 1993/02 y 1993/03 y la media móvil que se encuentra entre las observaciones 1993/03 y 1993/04 y el resultado se colocará (centrará) en la observación 1993/03. Observe que hasta antes de centrar el valor de la media móvil ésta no pertenecía, por así llamarle, a ninguna observación.

Así, la media móvil centrada para la observación 1993/03 es:

$$\bar{x}_1 = \frac{5871531.02 + 5912390.96}{2} = 5891960.99$$

Para la segunda media móvil centrada, ahora se omite el primer valor de la serie de la media móvil y se toman los siguientes dos, en este caso, las medias móviles que se encuentran entre las observaciones 1993/03 y 1993/04 y la media móvil entre las observaciones 1993/04 y 1994/01, obteniéndose la media móvil centrada para la observación 1993/04:

$$\bar{x}_2 = \frac{5912390.96 + 5996423.03}{2} = 5954406.99$$

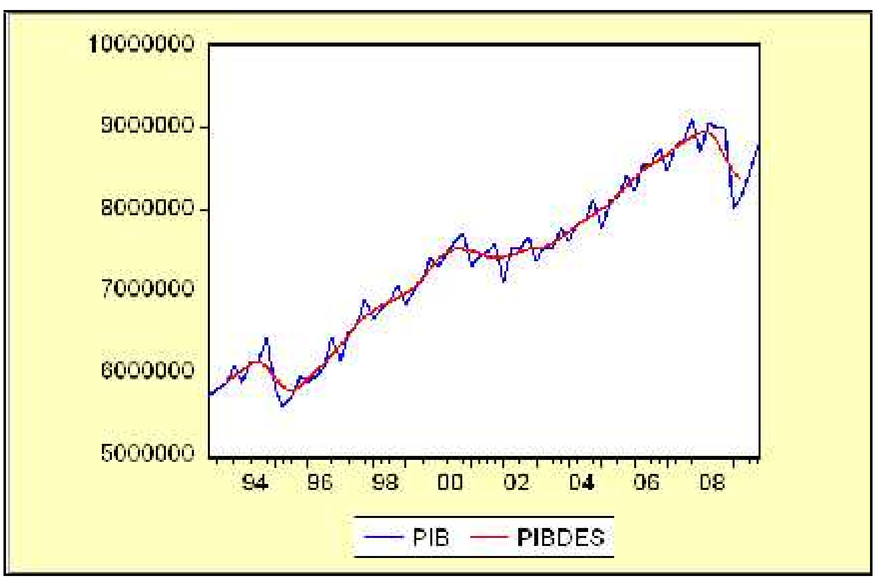
Para la tercera media móvil centrada se omiten los valores de las dos primeras medias móviles y se toman los siguientes dos, es decir, las medias móviles que se encuentran entre las observaciones 1993/04 y 1994/01 y la media móvil entre las observaciones 1994/01 y 1994/02, obteniéndose la media móvil centrada para la observación 1994/01:

$$\bar{x}_3 = \frac{5996423.03 + 6070233.1}{2} = 6033328.06$$

Y así sucesivamente hasta terminar la serie de las medias móviles centradas.

Posteriormente, como al trabajar en Excel se insertaron filas, ahora se copian los valores de la serie original del PIB y de la serie de las medias móviles centradas (serie desestacionalizada del PIB) y se pegan en una nueva hoja de Excel, únicamente que hay que pegarlos con formato especial. Así, una vez que se copiaron se da un click derecho y se selecciona la opción “pegado especial” y dentro de la ventana de opciones que se despliega en la ventana se selecciona la opción “valores”/Aceptar. Después, se eliminan las filas que separan los valores para poder copiarlos a Eviews siguiendo los pasos ya conocidos, en este caso se llevarán a Eviews las series del PIB (original, sin desestacionalizar) y la serie de las medias móviles centradas (serie del PIB desestacionalizada) y se graficarán de forma conjunta (tenga cuidado en pegar la serie desestacionalizada en la observación que corresponde, pues se perdieron 2 observaciones al inicio y 2 al final, por lo que la serie desestacionalizada parte del tercer trimestre de 1993 y la serie original del primer trimestre de 1993), obteniéndose:

Gráfica 10.4.2.1.2.1 Comportamiento de datos



Obsérvese como se suavizó o alisó el comportamiento de la serie del PIB (PIBDES), en este caso se eliminó el componente estacional de la serie original del PIB (de color azul) y para la serie PIBDES se perdieron 2 observaciones al inicio y 2 al final.

Con este método, como comenta Mason, et al.- (2000: 658) se identifica la tendencia a largo plazo en una serie de tiempo (ya que amortigua las fluctuaciones a corto plazo).

X.4.2.1.3.- Método de variaciones estacionales a partir de *índices* estacionales.

Este método consta de cinco pasos:

1. Se obtiene la serie de las medias móviles en caso de que m sea impar o la serie de las medias móviles centradas en caso de que m sea par.

2. Se calcula el *valor estacional específico* para cada observación dividiendo el valor de la serie original entre la media móvil (centrada o no). Dicho valor estacional específico expresa la razón del valor original de la serie de tiempo, al promedio móvil (centrado o no).

3. Los datos de los valores estacionales específicos se organizan en forma tabular y se calcula la media o promedio de cada frecuencia (de cada m) a través del tiempo, es decir, si m = 4, entonces al organizar la información en forma tabular se obtendrá el índice medio estacional para cada m, es decir, el índice medio estacional para todos los primeros, segundos, terceros y cuartos trimestres de todos los años que componen las serie de tiempo.

4. En teoría la suma de las medias de los índices estacionales debe ser igual a m, porque el promedio se fija en 1. Cuando la suma es distinta de m, entonces se aplica un factor de corrección a cada valor medio para obtener un total de m. El factor de corrección es:

$$Factor\ de\ corrección = \frac{m}{Suma\ \ de\ \ las\ \ medias\ \ de\ \ los\ \ índices\ \ estacionales}$$

Los índices obtenidos se deben expresar como porcentaje, por lo que deben de multiplicarse por 100.

Esta forma de promediar elimina la mayor parte de las fluctuaciones irregulares respecto de estacionales, y los índices que resultan indican el esquema representativo estacional de la serie, si es que presenta este componente.

5. Cada valor original de la serie se divide entre su respectivo índice medio estacional (ajustado o no) correspondiente, obteniéndose así la serie desestacionalizada. Con este método se recuperan las observaciones perdidas con el método anterior y es útil para realizar la predicción.

Ejemplo:

Retomando la serie de las medias móviles centradas para el ejemplo anterior (lo que equivale a realizar el paso 1 del método de índices estacionales), calculamos el valor estacional específico, en este caso para cada trimestre. Cabe mencionar que estos cálculos se continuarán realizando en Excel.

Así, para obtener el primer valor estacional específico dividimos el valor original del PIB en el tercer trimestre de 1993 con respecto a la media móvil centrada en dicho trimestre, obteniéndose:

$$Valor \quad estacional \quad específico = \frac{5857838.58}{5891960.99} = 0.99420865$$

El segundo valor estacional específico resulta de dividir el valor original del PIB en el cuarto trimestre de 1993 con respecto a la media móvil centrada en dicho trimestre, obteniéndose:

$$Valor \quad estacional \quad específico = \frac{6092879.1}{5954406.99} = 1.0232554$$

El tercer valor estacional específico resulta de dividir el valor original del PIB en el primer trimestre de 1994 con respecto a la media móvil centrada en dicho trimestre, obteniéndose:

$$Valor \quad estacional \quad específico = \frac{5896044.37}{6033328.06} = 0.97724578$$

Y así sucesivamente hasta calcular todos los valores específicos que componen la serie.

A continuación se muestra parte de los resultados del cálculo de los valores específicos:

Imagen 10.4.2.1.3.1 Obtención de Datos

Periodo	PIB	Media móvil	Media Móvil Centrada	Valores específicos
1993/Q1	5732604.63			
1993/Q2	5802801.77			
		5871531.02		
1993/Q3	5857838.59		5891960.99	0.99420865
		5912390.96		
1993/Q4	6092879.1		5954406.99	1.0232554
		5996423.03		
1994/Q1	5896044.37		6033328.06	0.97724573
		6070233.1		
1994/Q2	6138930.05		6111726.19	1.00445109
		6153219.27		
1994/Q3	6153078.88		6144433.21	1.00140707
		6135647.15		
1994/Q4	6424823.79		6068276.25	1.05875598
		6000905.34		
1995/Q1	5825755.88		5943259.7	0.98022905

Ahora organizamos de forma tabular los valores específicos, de tal suerte que se observan de la siguiente forma:

Tabla 10.4.2.1.3.1 Obtención de Datos

Año	Primer trimestre	Segundo trimestre	Tercer trimestre	Cuarto trimestre
1993			0.99420865	1.0232554
1994	0.97724578	1.00445109	1.00140707	1.05875598
1995	0.98022906	0.96090706	0.98508138	1.02256509
1996	0.99465799	0.98782873	0.99238171	1.03350732
1997	0.96914571	1.00359666	0.99655463	1.03084119
1998	0.98460726	0.99262392	1.00265877	1.02186121
1999	0.97883559	0.99183829	0.99897374	1.0189557
2000	0.98791675	0.99791	1.01095131	1.02572661
2001	0.97504713	0.99455995	1.0079805	1.02229852
2002	0.95732142	1.01074805	1.00395263	1.01975711
2003	0.97892295	0.99971917	0.99314096	1.0165615
2004	0.985488	1.00062956	0.99870576	1.02471625
2005	0.97087725	1.00509855	0.99622524	1.01666232
2006	0.98306533	1.0065597	0.99972958	1.01627042
2007	0.97751527	1.0026069	1.00271349	1.02249226
2008	0.97330259	1.01101827	1.01706785	1.03914835
2009	0.94673366	0.97230686		

Ahora procedemos a calcular el índice medio estacional para cada periodo, es decir, para todos los primeros, segundos, terceros y cuartos trimestres, de tal suerte que la suma de todos los valores específicos de los primeros trimestres es igual a 15.6209, de los segundos trimestres 15.9424, de los terceros trimestres 16.0017 y de los cuartos trimestres 16.4133.

Por Tanto la media de los índices estacionales por frecuencia (por m) son:

$$\ddot{x}_1 = \frac{15.6209118}{16} = 0.97630698$$

$$\ddot{x}_2 = \frac{15.9424027}{16} = 0.99640017$$

$$\ddot{x}_3 = \frac{16.0017333}{16} = 1.00010833$$

$$\bar{x}_4 = \frac{16.4133752}{16} = 1.02583595$$

Mismos que pueden visualizarse a continuación:

Tabla 10.4.2.1.3.2 Obtencion de Datos

Año	Primer trimestre	Segundo trimestre	Tercer trimestre	Cuarto trimestre
1993			0.99420865	1.0232554
1994	0.97724578	1.00445109	1.00140707	1.05875598
1995	0.98022906	0.96090706	0.98508138	1.02256509
1996	0.99465799	0.98782873	0.99238171	1.03350732
1997	0.96914571	1.00359666	0.99655463	1.03084119
1998	0.98460726	0.99262392	1.00265877	1.02186121
1999	0.97883559	0.99183829	0.99897374	1.0189557
2000	0.98791675	0.99791	1.01095131	1.02572661
2001	0.97504713	0.99455995	1.0079805	1.02229852
2002	0.95732142	1.01074805	1.00395263	1.01975711
2003	0.97892295	0.99971917	0.99314096	1.0165615
2004	0.985488	1.00062956	0.99870576	1.02471625
2005	0.97087725	1.00509855	0.99622524	1.01666232
2006	0.98306533	1.0065597	0.99972958	1.01627042
2007	0.97751527	1.0026069	1.00271349	1.02249226
2008	0.97330259	1.01101827	1.01706785	1.03914835
2009	0.94673366	0.97230686		
Sumatoria	15.6209118	15.9424027	16.0017333	16.4133752
n	16	16	16	16
Media	0.97630698	0.99640017	1.00010833	1.02583595

Ahora realizamos la suma de las medias de los índices estacionales:

$$Suma \quad de \quad índices \quad estacionales = 0.976306 + 0.996400 + 1.000108 + 1.025835 = 3.998651$$

En teoría la suma de estos índices debería ser de 4, pues m =4, por lo que tendremos que aplicar un factor de corrección:

$$Factor \quad de \quad corrección \quad = \frac{4}{3.998651} = 1.000337364$$

A continuación se multiplica cada índice medio estacional por el factor de corrección, obteniendo así el índice medio estacional ajustado para cada frecuencia:

$$Índice \quad ajustado_1 = 0.97630698 * 1.000337364 = 0.976636$$

$$Índice \quad ajustado_2 = 0.99640017 * 1.000337364 = 0.996736$$

$$Índice \quad ajustado_3 = 1.00010833 * 1.000337364 = 1.000445$$

$$Índice \quad ajustado_4 = 1.02583595 * 1.000337364 = 1.026182$$

Realizamos la suma de los índices estacionales ajustados y obtenemos:

$$Suma \quad de \quad índices \quad estacionales \quad ajustados = 0.976636 + 0.996736 + 1.000445 + 1.026182 = 4$$

En este caso se ajusto hacia arriba, ya que la suma anterior era menor que 4, en el caso de que la suma fuera mayor que 4, entonces, con el factor de corrección se ajustaría hacia abajo.

Ahora multiplicamos cada índice ajustado por 100, así, el índice medio ajustado del PIB del primer trimestre para todos los años es de 97.6636%, por lo que decimos que el PIB en el primer trimestre está 2.3364% (obtenido de restar 100-97.6636) por abajo del primer trimestre típico; el índice medio ajustado del PIB del segundo trimestre para todos los años es de 99.6736%, por lo que el PIB del segundo trimestre está 0.3264% por debajo del segundo trimestre típico; el índice medio ajustado del tercer trimestre del PIB para todos los años es de 100.0445%, por lo que el PIB en el tercer trimestre está 0.0445% por arriba del tercer trimestre típico o promedio; y finalmente el índice medio estacional ajustado del PIB en el cuarto trimestre es de 102.6182%, por lo que el PIB está 2.6182% por arriba del cuarto trimestre promedio. De aquí precisamente surge la necesidad de desestacionalizar la serie.

En el cuadro siguiente se muestran los índices medios estacionales y los índices medios estacionales ajustados:

Tabla 10.4.2.1.3.3 Obtencion de Datos

Año	Primer trimestre	Segundo trimestre	Tercer trimestre	Cuarto trimestre	
1993			0.99420865	1.0232554	
1994	0.97724578	1.00445109	1.00140707	1.05875598	
1995	0.98022906	0.96090706	0.98508138	1.02256509	
1996	0.99465799	0.98782873	0.99238171	1.03350732	
1997	0.96914571	1.00359666	0.99655463	1.03084119	
1998	0.98460726	0.99262392	1.00265877	1.02186121	
1999	0.97883559	0.99183829	0.99897374	1.0189557	
2000	0.98791675	0.99791	1.01095131	1.02572661	
2001	0.97504713	0.99455995	1.0079805	1.02229852	
2002	0.95732142	1.01074805	1.00395263	1.01975711	
2003	0.97892295	0.99971917	0.99314096	1.0165615	
2004	0.985488	1.00062956	0.99870576	1.02471625	
2005	0.97087725	1.00509855	0.99622524	1.01666232	
2006	0.98306533	1.0065597	0.99972958	1.01627042	
2007	0.97751527	1.0026069	1.00271349	1.02249226	
2008	0.97330259	1.01101827	1.01706785	1.03914835	
2009	0.94673366	0.97230686			
Sumatoria	15.6209118	15.9424027	16.0017333	16.4133752	
n	16	16	16	16	
Media	0.97630698	0.99640017	1.00010833	1.02583595	3.99865144
Índice ajustado	0.976636	0.996736	1.000445	1.026182	4.000000

Finalmente, ahora dividimos cada observación original de la serie del PIB entre su respectivo índice medio estacional ajustado, con lo cual obtendremos la serie del PIB desestacionalizada, es decir, todos los valores del PIB de los primeros trimestres se dividen entre el índice medio estacional ajustado del primer trimestre, en este caso, entre 0.976636; los valores de los segundos trimestres entre 0.996736; los de los terceros trimestres entre 1.000445; y los de los cuartos trimestres entre 1.026182.

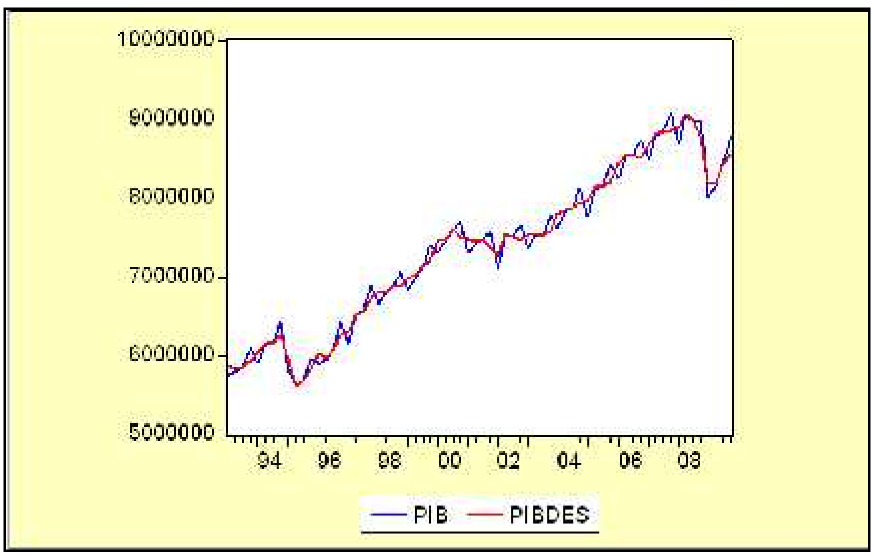
Así, por ejemplo, el valor del PIB en el primer trimestre de 1993 es de 5732604.63 este se divide entre 0.976636, obteniéndose el valor del primer trimestre de 1993 del PIB desestacionalizado; el valor del PIB en el segundo trimestre de 1993 es de 5802801.77 el cual se divide entre 0.996736, obteniéndose el valor del segundo trimestre de 1993 del PIB desestacionalizado; etc., ello puede observase a continuación en el cuadro siguiente:

Tabla 10.4.2.1.3.4 Obtención de Datos

Periodo	PIB	Índice ajustado	PIB desestacionalizado
1993/01	5732604.63	0.976636	5869745.37
1993/02	5802801.77	0.996736	5821804.14
1993/03	5857838.59	1.000445	5855233.01
1993/04	6092879.1	1.026182	5937425.43
1994/01	5896044.37	0.976636	6037095.06
1994/02	6138930.05	0.996736	6159033.13
1994/03	6153078.88	1.000445	6150341.98
1994/04	6424823.79	1.026182	6260900.88
1995/01	5825755.88	0.976636	5965125.07
1995/02	5599962.82	0.996736	5618300.95
1995/03	5691913.71	1.000445	5689381.94
1995/04	5962216.99	1.026182	5810097.03

Ahora, copiamos la serie original del PIB y la del PIB desestacionalizada a Eviews y las graficamos conjuntamente, obteniendo el siguiente cuadro:

Gráfico 10.4.2.1.3.1 Series.



La serie original del PIB está de color azul y la serie del PIB desestacionalizada con el método de variaciones estacionales a partir de índices estacionales de color rojo. Como puede apreciarse con este método se recuperan las observaciones perdidas y se desestacionaliza la serie, en este caso, los primeros y segundos trimestres de cada año se ajustaron hacia arriba y los terceros y cuartos hacia abajo, eliminando así el componente estacional. Asimismo, con esta serie desestacionalizada del PIB es posible realizar predicción.

X.4.2.1.4.-Método de las diferencias estacionales.

Con este método se elimina gran parte del impacto estacional de la serie temporal. Para ello (como se hizo antes cuando se ilustró este concepto de diferenciación) se obtiene la serie de diferencias de orden m (que es el periodo estacional definido), definiendo cada diferencia como $Z_t = Y_t - Y_{t-m}$.

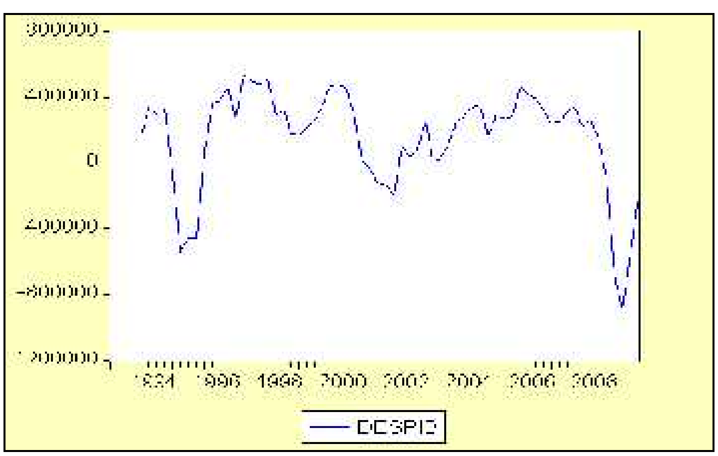
Obsérvese que con este procedimiento de diferenciación de orden m se pierden m observaciones de la serie inicial u original.

Ejemplo:

Continuando con la serie del PIB de forma trimestral desde el primer trimestre de 1993 hasta el cuarto trimestre de 2009 y dado que anteriormente se había detectado un claro comportamiento estacional en ella, ahora se procede a desestacionalizarla en Eviews con el método de las diferencias estacionales.

Así, una vez que la serie original se copió a Eviews, en este caso el PIB, en la barra de comandos escribimos: “genr despib=pib-pib(-4)” donde el 4 indica que la serie se diferenciará de la siguiente manera: al primer trimestre de 1994 se le resta el primer trimestre de 1993, al segundo semestre de 1994 se le resta el segundo de 1993, al tercer trimestre de 1994 se le resta el tercer trimestre de 1993, al cuarto trimestre de 1994 se le resta el cuarto de 1993, al primer trimestre de 1995 se le resta el primer trimestre de 1994, etc., perdiéndose por tanto 4 observaciones ya que la frecuencia en este caso es trimestral, es decir, m = 4. Una vez obtenida la serie “despib” la abrimos y graficamos con el método habitual y obtenemos:

Gráfico 10.4.2.1.4.1 Comportamiento de la serie



Observe que con este método se eliminó tanto la estacionalidad como la tendencia, así, la serie "despib" obtenida mediante el método de diferencias estacionales permite eliminar al menos dos componentes de la serie, sin embargo, se pierden m observaciones.

X.5.- Predicción de los valores de Yt.

Una vez que se ha identificado la existencia o tipo de variaciones (tendencia, estacional, cíclica y residual) que puede presentar Y_t, y que dichas oscilaciones se hayan eliminado con cualesquiera de los métodos antes señalados, ahora se está en condiciones de hacer predicciones de los valores de Y_t.

Al respecto, las predicciones pueden ser de dos tipos: *condicionales o incondicionales*.

a).- Las condicionales tienen como fundamento la relación causal que existe entre la variable explicativa Y_t y su explicativa X_t, de manera que Y_t se puede predecir sí y sólo sí se conocen los valores futuros de X_t.

b).- Las incondicionales se realizan usando métodos autoproyectivos que, como su nombre lo indica, para predecir los valores de Y_t sólo se necesita conocer sus valores pasados y el actual, por lo que también se conocen como autorregresivos, AR, con los que se hallarán sus valores futuros; en otras palabras, no es necesario conocer X_t. Estos métodos pueden basarse en dos enfoques alternativos:

1. El determinista o clásico; y
2. El estocástico o moderno basado en Box and Jenkins.

Se recomienda el determinista cuando la muestra es pequeña; el estocástico cuando la muestra es grande.

Derivado de las recomendaciones anteriores es que decimos que según sea la predicción es que se seguirán ciertos métodos, que son más apropiados para predecir. Ejemplo:

1. Para el largo plazo se sugiere el análisis de tendencia;
2. Para el corto y mediano plazo los métodos econométricos.

Al respecto, para la proyección a corto plazo se deben de tener en cuenta las variaciones estacionales, E_t. Para el mediano plazo, las oscilaciones cíclicas.

Ejemplo:

Retomando la serie del PIB desestacionalizada con el método de variaciones estacionales a partir de índices estacionales realizaremos la predicción para los siguientes cuatro trimestres de 2010. Para ello, dado que la tendencia de la serie del PIB es creciente lineal se correrá la regresión de la serie desestacionalizada frente al tiempo, de tal suerte que:

PIB desestacionalizado = a + b (tiempo)

En este caso la variable tiempo tomara los valores 1, 2, 3,…, 68 para 1993/01, 1993/02,…, 2009/04 respectivamente.

Así, en Eviews generamos la serie tiempo escribiendo en la barra de comandos "genr tiempo=@trend+1" y luego corremos la regresión de la serie del PIB desestacionalizada (para ello no olvide que ya debe tener copiada su serie desestacionalizada), para ello escribimos en la barra de comandos "ls despib c tiempo" con lo que obtenemos la siguiente ecuación de regresión:

Cuadro 10.5.1 Regresión

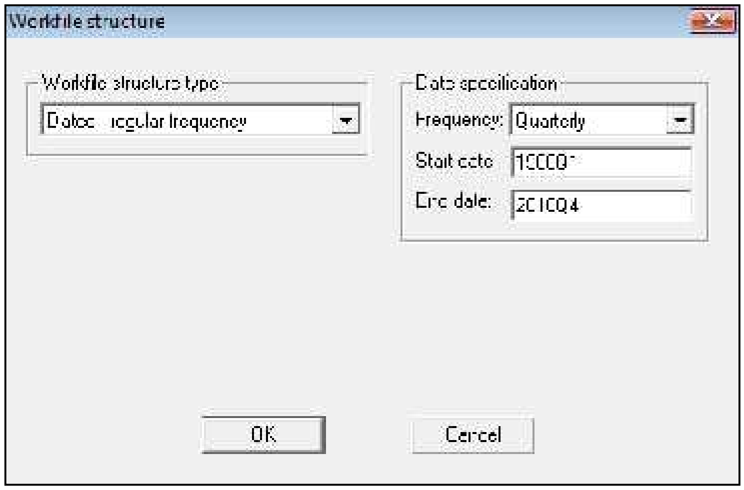
Dependent Variable: PIBDES
Method: Least Squares
Date: 11/21/10 Time: 22:59
Sample: 1993Q1 2009Q4
Included observations: 68

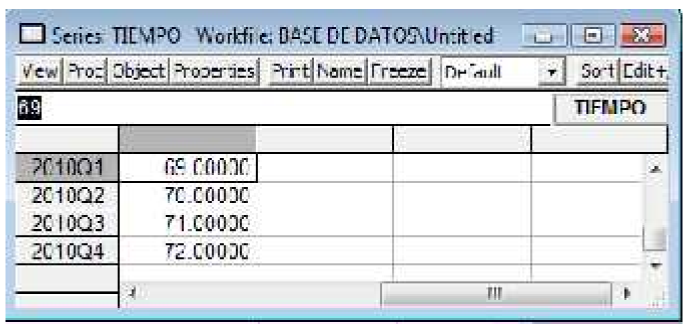
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5661140.	63099.85	89.71715	0.0000
TIEMPO	40047.21	1309.715	30.72700	0.0000

R-squared	0.934663	Mean dependent var	7346360
Adjusted R-squared	0.933673	S.D. dependent var	999000.7
S.E. of regression	257303.1	Akaike info criterion	27.70207
Sum squared resid	4.37E+12	Schwarz criterion	27.74015
Log likelihood	-942.0175	F-statistic	944.1503
Durbin-Watson stat	0.292210	Prob(F-statistic)	0.000000

Como puede observarse los coeficientes son estadísticamente significativos y el R2 es alto.

Ahora, guardamos la ecuación de regresión pulsando el botón "name" dentro de la ventana de la regresión y conservamos el nombre que el programa le asigna por default "eq01"; vamos a la ventana del workfile y damos doble click sobre la leyenda "Range" en la ventana que nos despliega en "End date" escribimos 2010Q4/Ok., con lo cual también se incrementará el "Sample" en cuatro observaciones (esto siempre y cuando trabaje con Eviews 5), abrimos la serie tiempo y en la observación 2010Q1 ponemos 69, en la 2010Q2 escribimos 70, en la 2010Q3 escribimos 71 y en la 2010Q4 escribimos 72.





Regresamos a la ventana de la ecuación de regresión llamada “eq01” y oprimimos el botón de “Forecast” y en la ventana que nos despliega en “Forecast name” preservamos el nombre que por default otorga el programa a la serie “DESPIBF”, verificamos que el “Forecast Sample” vaya de 1993Q1 a 2010Q4 pues nos interesa predecir el valor de los siguientes cuatro trimestres, Ok.

En la ventana del workfile aparecerá la nueva serie, seleccionamos la serie PIBDES y PIBDESF y las abrimos como grupo y observamos los siguientes valores para el PIB estimados a partir de la ecuación de regresión anterior:

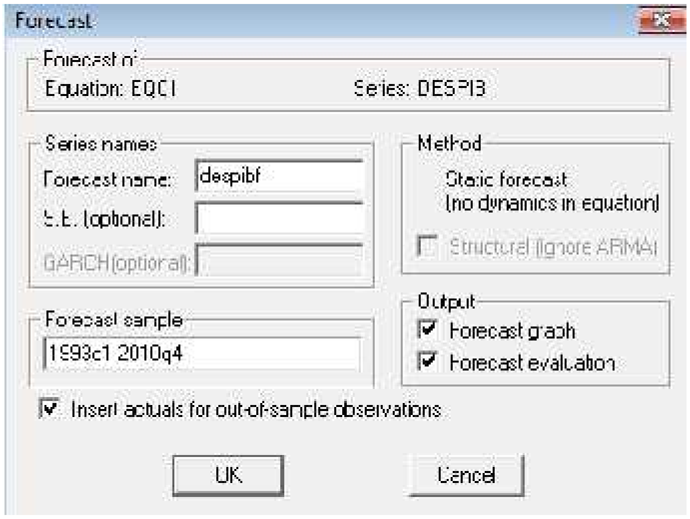
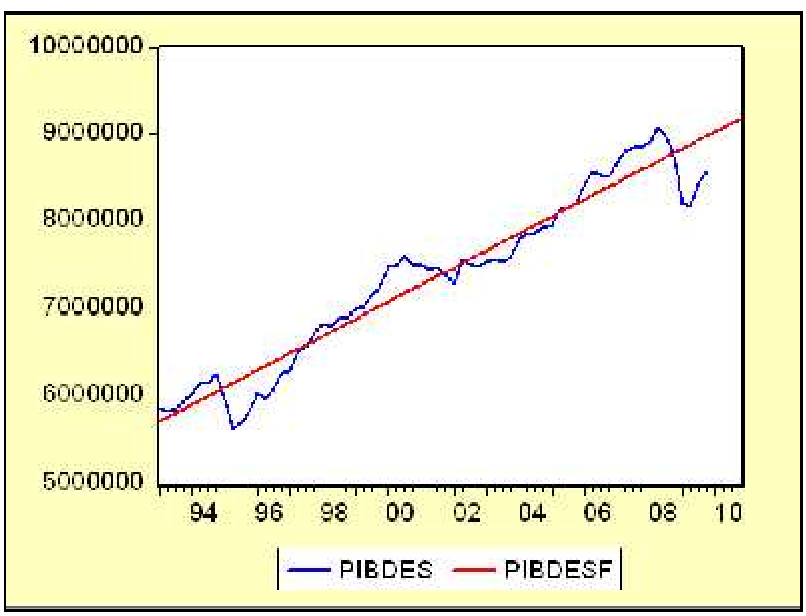


Tabla 10.5.1 Datos

Periodo	PIB desestacionalizado	Estimación del PIB desestacionalizado
2008Q4	8755017.75	8787361.215
2009Q1	8205694.68	8836208.426
2009Q2	8165970.48	8885055.637
2009Q3	8445231.91	8933902.848
2009Q4	8556289.32	8982750.059
2010Q1		9031597.269
2010Q2		9080444.48
2010Q3		9129291.691
2010Q4		9178138.902

Graficamos las dos series mediante los pasos habituales y obtenemos:

Gráfico 10.5.1 Serie



Observe que la tendencia de la serie del PIB desestacionalizada estimada (PIBDESF) siempre es creciente, mientras que la serie del PIB desestacionalizada (PIBDES) no lo es, por tanto los valores estimados del PIB para los próximos 4 trimestres de 2010 deberán ajustarse a partir de multiplicar sus valores estimados (obtenidos a partir de la ecuación de regresión anterior) por sus respectivos índices medios estacionales ajustados que se calcularon previamente, así los valores ajustados de la predicción del PIB desestacionalizado serán:

Para el primer trimestre de 2010:

PIBDES2010Q1 = 9031597.269*0.976636 = 8820583.03

Para el segundo trimestre de 2010:

PIBDES2010Q2 = 9080444.48*0.996736 = 9050805.91

Para el tercer trimestre de 2010:

PIBDES2010Q3 = 9129291.691*1.000445 = 9133354.23

Para el cuarto trimestre de 2010:

PIBDES2010Q4 = 9178138.902*1.026182 = 9418440.93
Los resultados se muestran en el siguiente tabla:

Tabla 10.5.2 Datos

Periodo	PIB desestacionalizado	Estimación del PIB desestacionalizado	Índice ajustado	Estimación del PIB
2008Q4	8755017.75	8787361.215		
2009Q1	8205694.68	8836208.426		
2009Q2	8165970.48	8885055.637		
2009Q3	8445231.91	8933902.848		
2009Q4	8556289.32	8982750.059		
2010Q1		9031597.269	0.976636	8820583.03
2010Q2		9080444.48	0.996736	9050805.91
2010Q3		9129291.691	1.000445	9133354.23
2010Q4		9178138.902	1.026182	9418440.93

Se observa que al multiplicar los valores estimados mediante la ecuación de regresión por el índice medio estacional ajustado la estimación del PIB para los próximos cuatro trimestres ya no posee un comportamiento estacional.

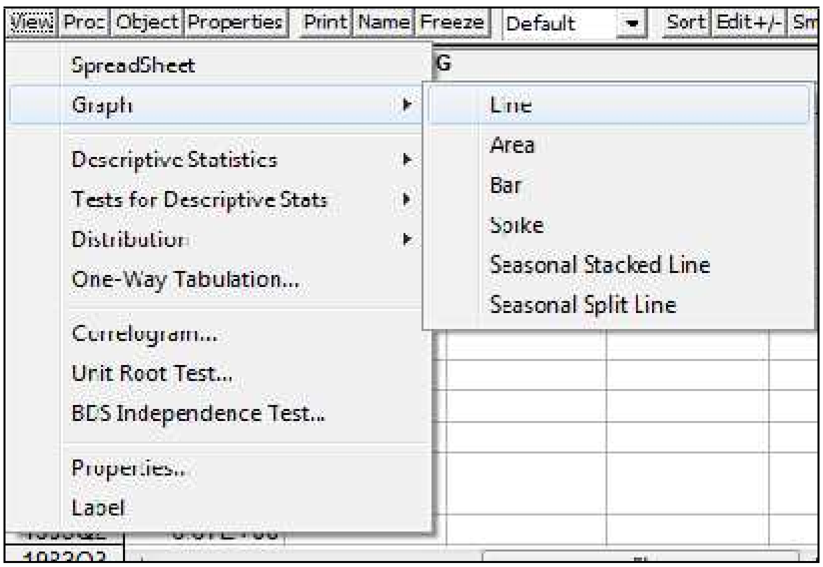
Ejemplo 4

Descomposición de series de tiempo con e views 5.

En este ejemplo se descompondrá la serie de Consumo Gubernamental (CG), la cual cuenta con 129 observaciones, que van desde el primer trimestre de 1980 hasta el primer trimestre de 2012.La finalidad es conocer los componentes de una serie de tiempo.

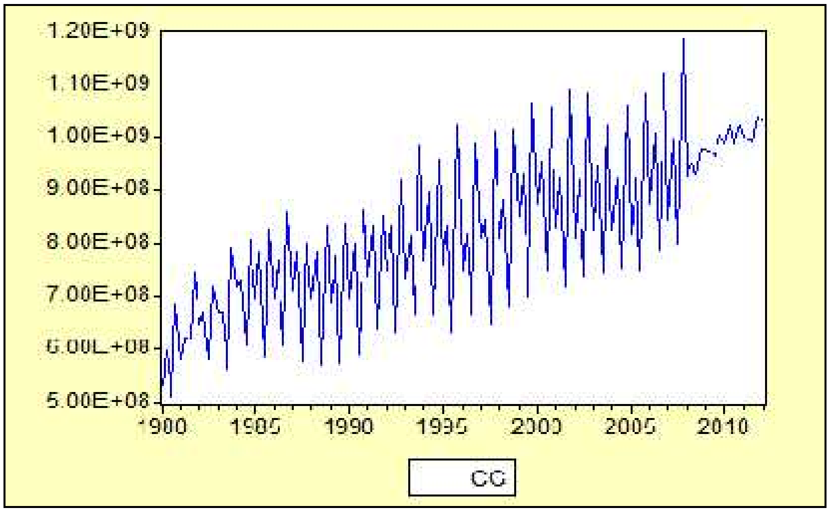
Como se ha venido haciendo en este libro se introducirá al workfile de Eviews 5 la serie del CG. Posteriormente se abrirá la serie y en la opción view/graph/line.

Cuadro10.5.2. Descripcion



Del mecanismo anterior se obtiene la gráfica que nos indica el comportamiento de la serie, a partir de dicha serie se pueden hacer algunas conjeturas, que se comprobarán con pruebas posteriores.

Gráfico 10.5.2 Comportamiento de la serie



Inicialmente, y como se puede observar en ese Cuadro la serie correspondiente al CG presenta una tendencia ascendente, el ciclo y la estacionalidad se definen con claridad en este paso, y la aleatoriedad es constante a lo largo de la serie.

Con el fin de complementar la información del ejercicio anterior (ejercicio 1), se trabajarán dos métodos de descomposición adicionales a los ya presentados.

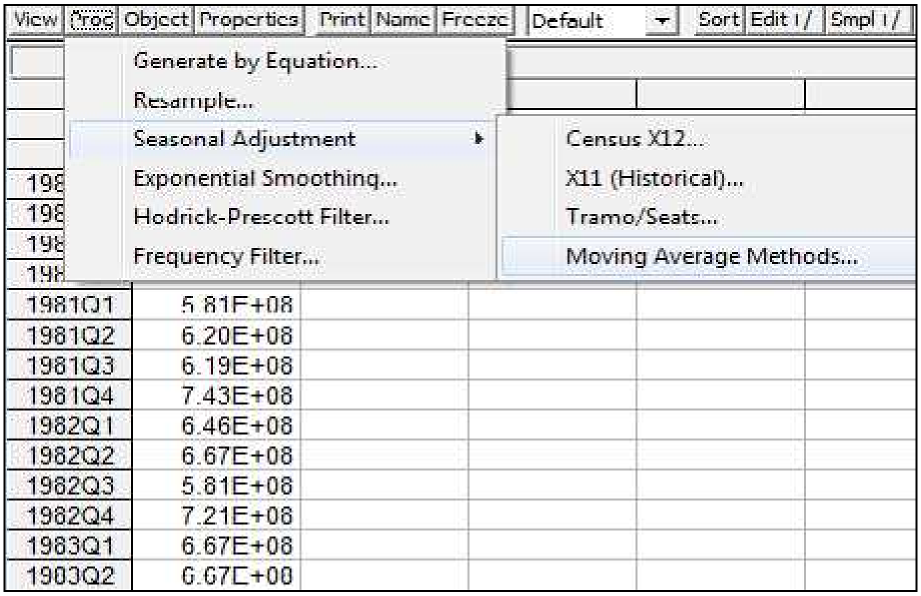
Método aditivo.

Este método se caracteriza por sumar los cuatro elementos que conforman una serie de tiempo: Tendencia (T) + Ciclo (C)+ Estacionalidad (Sn) + Aleatoriedad (A). Este se utiliza en caso de que la variación sea constante.

Componente Estacional

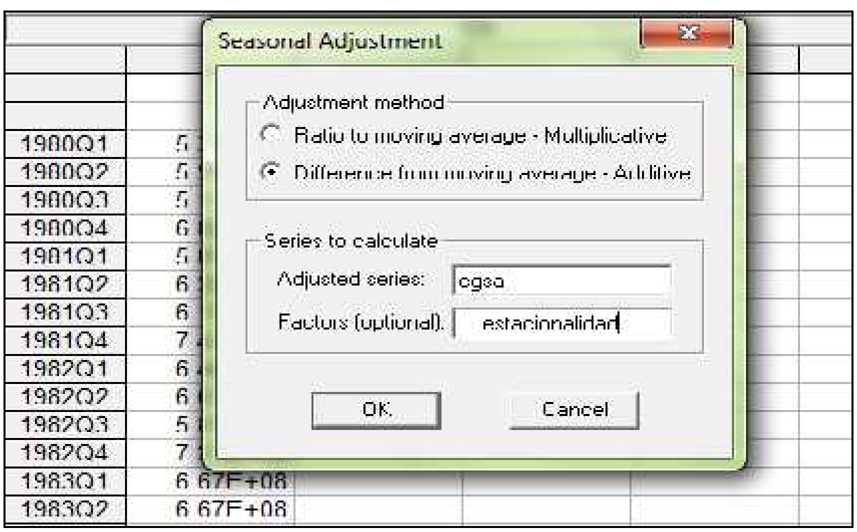
Para comenzar el análisis se abre la serie CG, en la barra de menú se selecciona Proc/ Seasional Adjustment / Moving Average. Tal como se muestra en el cuadro siguiente.

Cuadro 10.5.3 Descripción



Después de haber seleccionado dicha opción, se abre una nueva pantalla de Seasional Adjustment, en él aparece la opción Adjustment method/Difference from moving average- Additive. En la siguiente opción: Series to calcúlate aparece el nombre de la serie creada como cgsa por default, se puede cambiar si así se desea, en Factors (optional) se sugiere incluir el nombre del componente, en este caso: estacionalidad.

Cuadro 10.5.4 Descripcion



En este paso la serie genera medias móviles con el fin de eliminar el componente estacional de la serie.

A continuación Eviews 5 arroja el resultado de haber estimado las medias móviles.

Cuadro 10.5.5 Descripción

ESTACIONALIDAD
Sample: 1980Q1 2012Q1
Included observations: 129
Ratio to Moving Average
Original Series: CG
Adjusted Series: CGSA

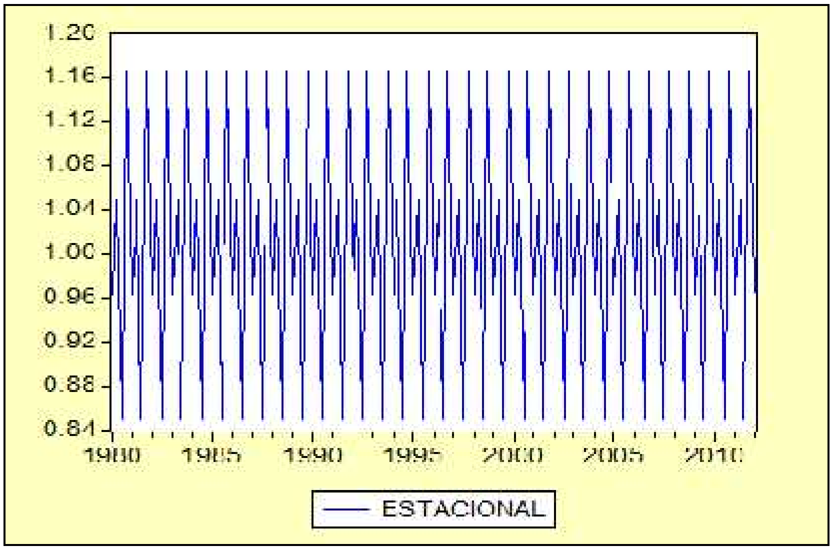
Scaling Factors:

1	0.963332
2	1.048273
3	0.849240
4	1.166056

Eviews proporciona los datos en trimestres e indican la estacionalidad de la serie del CG.

Como se aprecia en el gráfico correspondiente a la estacionalidad, Cuadro 10.5.5 la serie fluctúa entre 1.1660 y 0.8492 y se han eliminado los demás componentes. En este caso se está trabajando con trimestres y se observa que el movimiento que se da en el primer trimestre es parecido al que se da en el tercero, mientras que el movimiento que se da en el segundo trimestre es parecido al que se da en el cuarto trimestre; por lo tanto, se puede decir que hay cierta regularidad.

Cuadro 10.5.6 Descripción

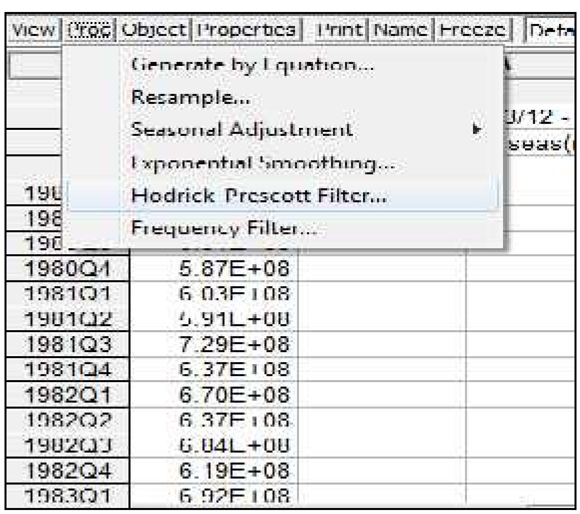


Componente Tendencia.

En esta parte de la descomposición, se utiliza el filtro Hodrick Prescott, cuya principal función es linealizar los parámetros. El parámetro, que en este caso es λ, modula la tendencia y, de acuerdo con el valor del parámetro, será la longitud del ciclo y la periodicidad de los datos. Por lo tanto, el Filtro Hodrick Prescott va a descomponer la serie en ciclo y tendencia.

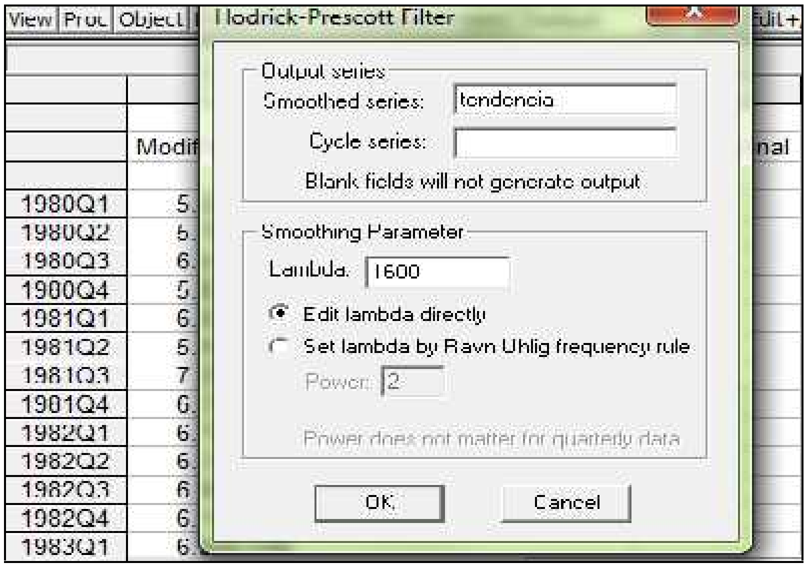
Retomando el ejemplo, es necesario abrir la serie que creamos bajo el nombre de CGSA mediante el Seasonal Adjustment, en la barra de menú seleccionar Proc/HP Filter, como se muestra a continuación.

Cuadro 10.5.7 Descripción



Enseguida aparece un cuadro de diálogo, en Output series viene la opción Smoothed series: Se sugiere cambiar el nombre predeterminado por “tendencia”, de esa forma se crea el segundo componente de la serie. A continuación se ilustra lo anterior.

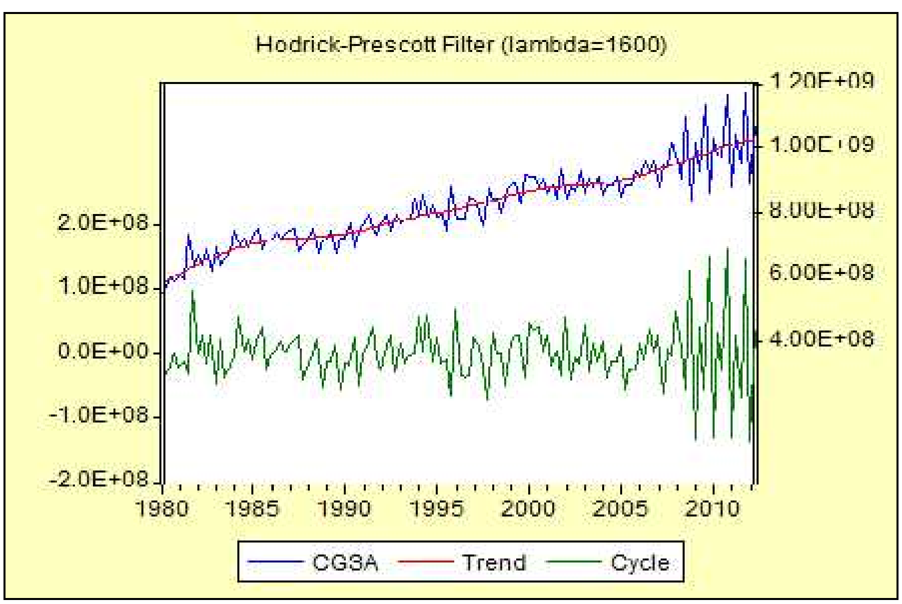
Cuadro 10.5.8 Descripción



Como se mencionó con antelación, el parámetro (λ) es muy importante, es por ello que si no se tiene un cálculo preciso de la estimación de éste se deje el valor que trae predeterminado, que en este caso es de 1600.

En cuanto se da click en la opción ok del cuadro de diálogo, aparece un gráfico nuevo.

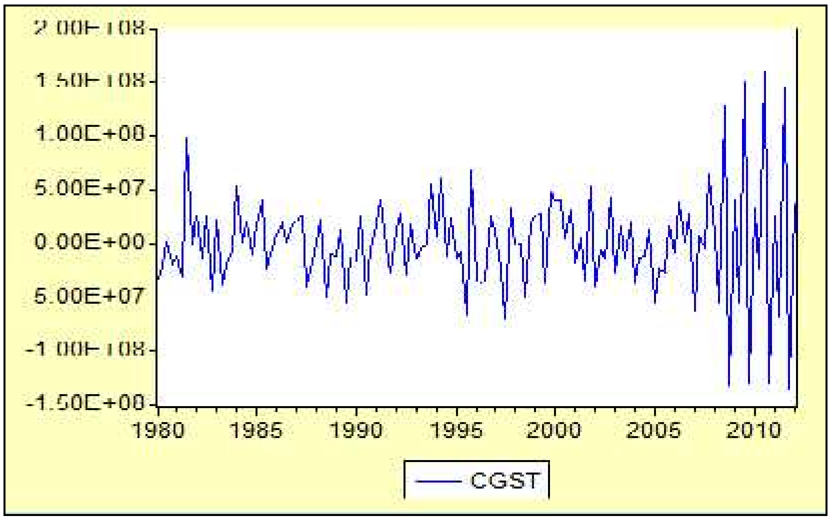
Gráfico 10.5.2 Comportamiento de series



En este cuadro se muestra la tendencia en rojo y la estacionalidad en azul, el ciclo se aprecia en verde, sin embargo hace falta llevar a cabo algunos paso más para definirlos de manera correcta.

Así, En el cuadro de texto de Eviews 5 se escribe: genr cgst=cgsa-tendencia, lo que devolverá a la serie sin estacionalidad y sin tendencia, como se muestra en el siguiente cuadro.

Gráfico 10.5.3 Comportamiento de series

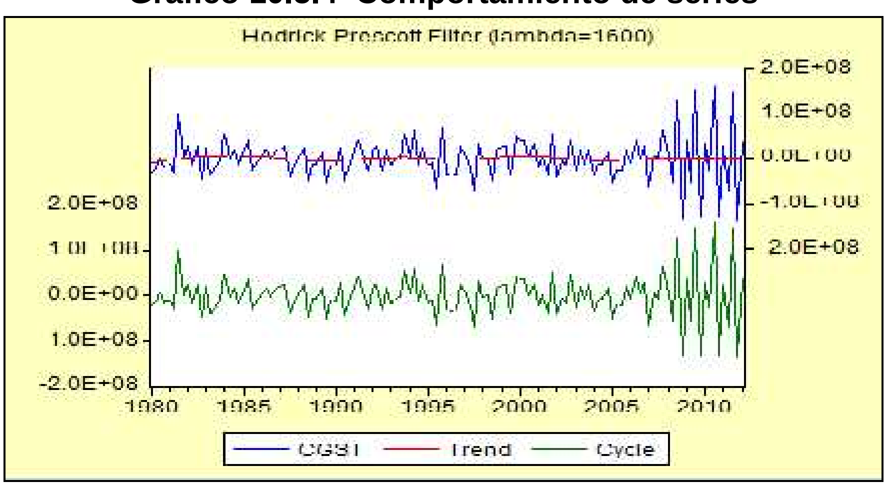


Este paso es el preámbulo del componente ciclo, aun falta hacer algunos pasos más.

Componente ciclo.

Se abre la serie generada: cgst/procs/hp filter, a continuación aparece una cuadro de diálogo, lo único que se modifica ahí es Smoothed series: Ciclo.

Gráfico 10.5.4 Comportamiento de series



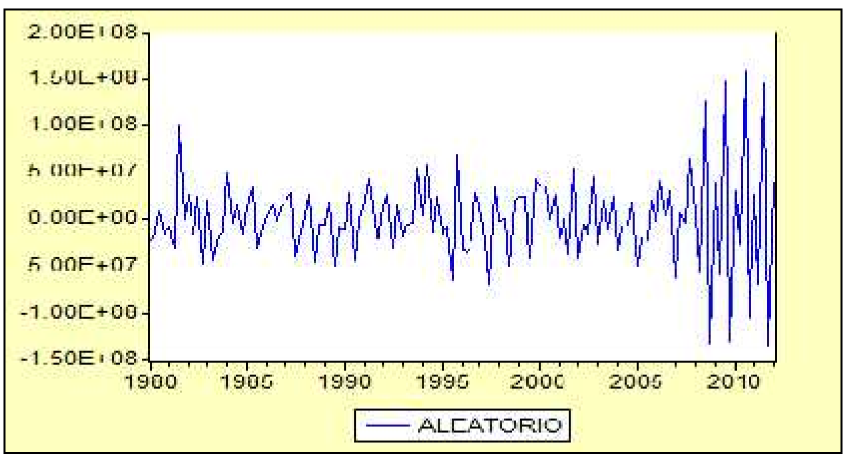
Aquí se puede apreciar el componente ciclo, el cual no es diferente de las estimaciones que se habían venido haciendo, lo importante de este gráfico es la línea verde, la cual presenta ciclos muy irregulares, en los que nos se aprecia con claridad el periodo de auge, expansión, desaceleración, recesión y depresión que se presenta regularmente. Existe la opinión generalizada de que a partir de 2008 el CG ha tenido un comportamiento muy volátil, más que el registrada desde 1980.

Componente Aleatorio.

En este caso se espera que la esperanza matemática (media de los errores) sea igual a cero, es decir. $E(u) = 0$ para que los valores no se volatilicen y la serie se comporte de la manera más lineal posible.

En el cuadro de texto de Eviews 5: genr aleatorio=cgst-ciclo, se abre la serie, view/graph/line. El gráfico resultante se presenta a continuación.

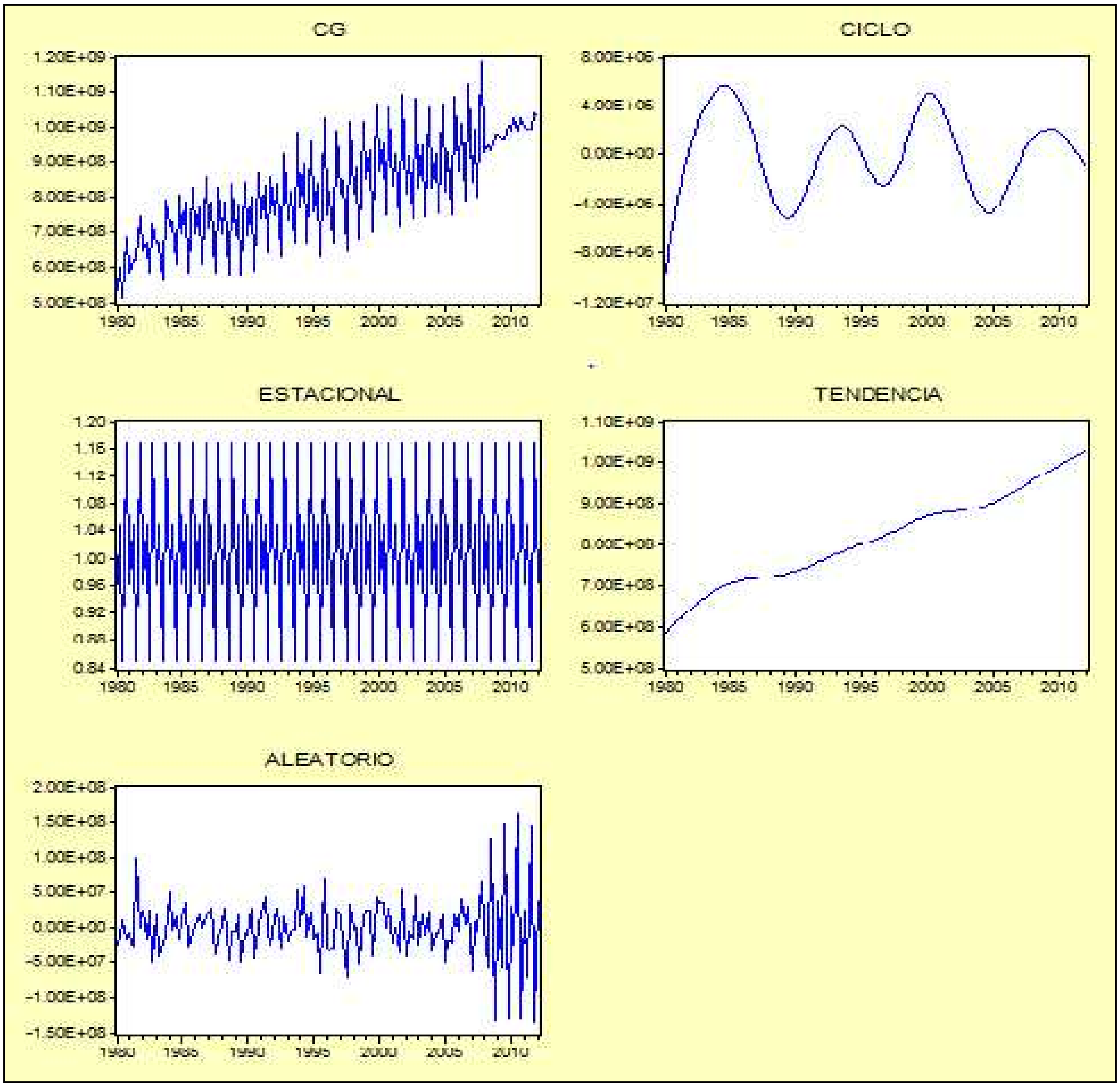
Gráfico 10.5.5 Comportamiento de series



Derivado de lo anterior se puede decir que es el comportamiento de toda la serie es aleatorio, porque después de haber realizado diversas pruebas y filtros el comportamiento de la serie no cambió.

Para obtener todas las gráficas de este método y, a manera de resumen, se seleccionan en el workfile cg, estacional, tendencia, ciclo y aleatorio, con click derecho, open/as group, se abre la serie, en la opción view/multiple graph/line. El resultado de dicha operación es el siguiente:

Gráfico 10.5.6 Comportamiento de series



En este Cuadro se puede apreciar claramente la descomposición de la serie en todos sus componentes.

En el primer recuadro (CG) se aprecia el comportamiento de la serie sin ningún filtro o método de suavizamiento.

En el segundo recuadro (ciclo) se distinguen cuatro ciclos, el primer auge tiene lugar en 1985, año en el que la Ciudad de México sufrió el sismo más devastador de la historia del país, razón por la cual el gobierno pudo haber tenido un consumo mayor. Por otra parte, el segundo auge se encuentra en 1993, año en que las variables macroeconómicas mostraban estabilidad y congruencia en su comportamiento tal que daban confianza para la futura evolución de la economía. En este contexto es que un año después se firmó el Tratado de Libre Comercio. El tercer auge en el ciclo se presenta en 2001, año de la crisis financiera provocada por el acelerado crecimiento de las empresas .com, debido a esto el gobierno tuvo que consumir más para evitar la profundización de la crisis.

El cuarto y último auge se presentó en 2009, año en el que México, además de encontrarse inmerso en la crisis económica mundial, sufrió una desaceleración debido a la epidemia de influenza que tuvo lugar en el verano de dicho año. En ese año se incrementó el consumo gubernamental para evitar la caída del producto, el esfuerzo fue en vano porque el PIB decreció.

En el tercer recuadro se observa el componente estacional, cuya incidencia se da a aproximadamente en los meses de junio-julio de cada año.

En el cuarto recuadro se presenta la tendencia, la cual es ascendente y constante.

El quinto y último recuadro presenta el componente aleatorio, el cual tiene como característica principal la volatilidad de 2008 a 2012 debido al comportamiento de los mercados a nivel mundial, los cuales son impredecibles y no controlables.

Comentarios:

Con este instrumental econométrico es posible diagnosticar las fluctuaciones de la economía y por consiguiente, visualizar y aplicar oportunamente medidas correctivas que las elimine o reduzcan con el objetivo principal de lograr un crecimiento sostenido y sustentable a partir de la estabilidad que se logre con ellas.

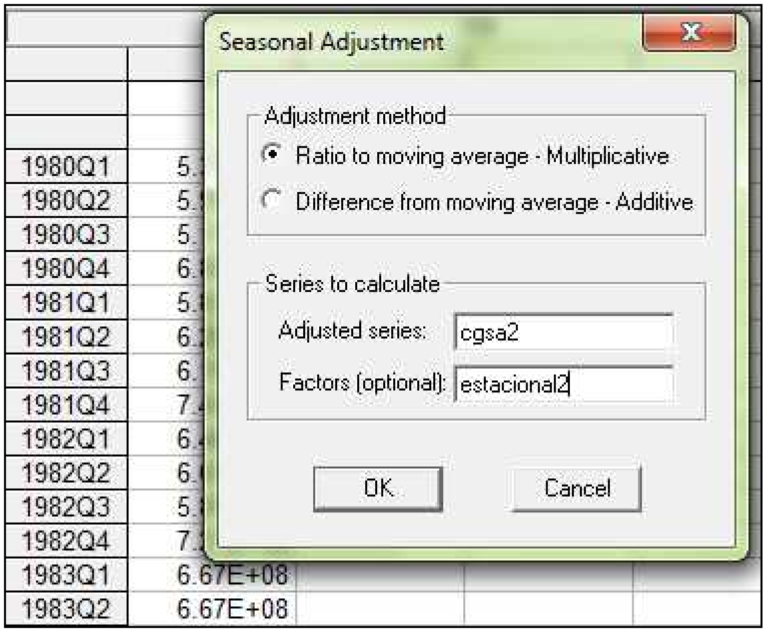
Método multiplicativo.

Este método se caracteriza por multiplicar los cuatro elementos que conforman una serie de tiempo: Tendencia (T) * Ciclo (C)* Estacionalidad (Sn) * Aleatoriedad (A). Este se utiliza en caso de que se haya observado que la variación sea creciente. La finalidad, como ya se mencionó, es descomponer la serie en sus cuatro componentes.

Componente Estacional.

Para empezar se abre la serie CG, que es con la que se ha venido trabajado. En la barra de menú se selecciona: Procs/Seasonal Adjustment en el cuadro de diálogo que aparece selecconar: Adjustment method/Ratio to moving average-Multiplicative. En el cuadro de diálogo Series to Calculate/Adjusted series, agregar un 2 al nombre predeterminado, es decir: cgsa2, esto debido a que seguramente se sigue trabajando con la misma base de datos que se utilizó en el método aditivo y el “2” sirve para diferenciar. En Factors(optional): escribir estacional2, como se muestra en el siguiente cuadro.

Cuadro 10.5.3 Descripción



Al dar click en ok al cuadro de diálogo anterior, aparecen los resultados del componente estacional trimestral presentado en el cuadro siguiente.

Cuadro 10.5.9 Componente estacional

ESTACIONAL 2
Sample: 1980Q1 2012Q1
Included observations: 129
Ratio to Moving Average
Original Series: CG
Adjusted Series: CGSA2

Scaling Factors:

1	0.963332
2	1.048273
3	0.849240
4	1.166056

El rango de estacionalidad oscila entre 0.8492 y 1.1660 para cada trimestre de 1980 a 2012. El primer y tercer trimestre, tienen un comportamiento semejante, lo mismo que el segundo y cuarto trimestre. Lo cual da indicios del movimiento de la serie.

Componente Tendencia.

Para poder obtener el componente tendencia es necesario generar la variable, entonces se escribe en el cuadro de texto de Eviews 5: genr t=@trend+1.

Después de haber generado la variable se realiza una regresión, se escribe en el cuadro de texto de Eviews 5: LS CG C T. A continuación se presenta el resultado de dicha regresión.

Cuadro 10.5.10 Regresión

Dependent Variable: CG
Method: Least Squares
TENDENCIA 2
Sample: 1980Q1 2012Q1
Included observations: 129

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.25E+08	18461466	33.87521	0.0000
T	2981694.	246445.1	12.09881	0.0000

R-squared	0.535447	Mean dependent var	8.19E+08
Adjusted R-squared	0.531789	S.D. dependent var	1.52E+08
S.E. of regression	1.04E+08	Akaike info criterion	39.77752
Sum squared resid	1.38E+18	Schwarz criterion	39.82186
Log likelihood	-2563.650	F-statistic	146.3813
Durbin-Watson stat	3.326056	Prob(F-statistic)	0.000000

Del cuadro anterior, se observa que la variables no son estadísticamente significativas, es decir el valor de sus parámetros es de cero, aunado a eso, la bondad de ajuste global (R2) de la regresión es bajo, con un 53.54%, por lo que se recomienda realizar nuevamente la regresión pero ahora con la variable tendencia elevada al cuadrado con el fin de que exista un mayor ajuste global de la prueba. Para ello se escribe en el cuadro de texto de Eviews 5: LS CG C T Tˆ 2. El resultado de la regresión se presenta a continuación.

Cuadro 10.5.11 Regresión

Dependent Variable: CG
Method: Least Squares
TENDENCIA ^2
Sample: 1980Q1 2012Q1
Included observations: 129

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.20E+08	28067211	22.08921	0.0000
T	3229185.	996731.5	3.239775	0.0015
T^2	-1903.780	7427.314	-0.256321	0.7981

R squared	0.535589	Mean dependent var	8.19E+08
Adjusted R-squared	0.528319	S.D. dependent var	1.52E+08
S.E. of regression	1.05E+08	Akaike info criterion	39.79250
Sum squared resid	1.38E+18	Schwarz criterion	39.85901
Log likelihood	-2563.516	F-statistic	72.68506
Durbin-Watson stat	3.327764	Prob(F-statistic)	0.000000

En este caso nuevamente la constante y la tendencia son significativas, sin embargo, la tendencia aumentó su probabilidad con respecto a su valor del cuadro pasado; mientras que la tendencia elevada al cuadrado resulta no significativa. El ajuste global de la regresión es sólo 2% mayor que en la regresión pasada, lo cual nuevamente indica un ajuste bajo entre variables.

Con el fin de lograr un mejor ajuste, se realizó una regresión con la variable tendencia, tendencia elevada el cuadrado y al cubo. Los resultados se observan a continuación.

Cuadro 10.5.12 Regresión

Dependent Variable: CG
Method: Least Squares
TENDENCIA ^3
Sample: 1980Q1 2012Q1
Included observations: 129

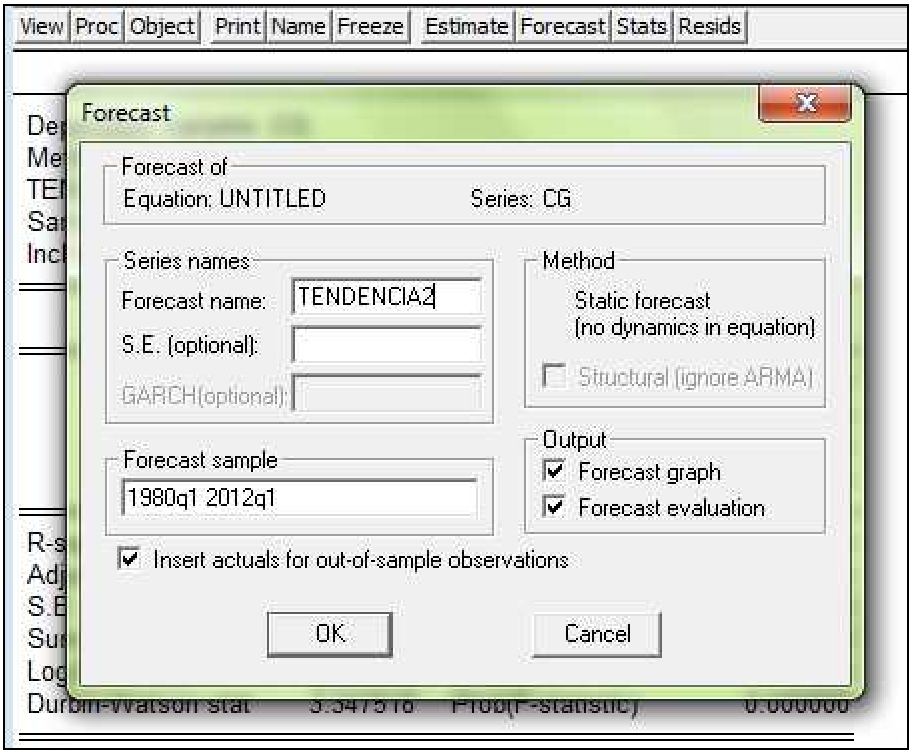
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.98E+08	37979126	15.74063	0.0000
T	5236725.	2520316.	2.077805	0.0398
T^2	-40361.63	44955.79	-0.897807	0.3710
T^3	197.2198	227.3681	0.867403	0.3874

R-squared	0.538467	Mean dependent var	8.19E+08
Adjusted R-squared	0.527391	S.D. dependent var	1.52E+08
S.E. of regression	1.05E+08	Akaike info criterion	39.80200
Sum squared resid	1.37E+18	Schwarz criterion	39.89068
Log likelihood	-2563.229	F-statistic	48.61228
Durbin-Watson stat	3.347518	Prob(F-statistic)	0.000000

En este caso las variables tendencia (tendencia, tendencia elevada el cuadrado y al cubo) son significativas, sin embargo, la constante sigue siendo no significativa. El ajuste global se incrementó poco, y con 53.84%, se vuelve a concluir que las variables tendencia, apenas explican la mitad del comportamiento del CG.

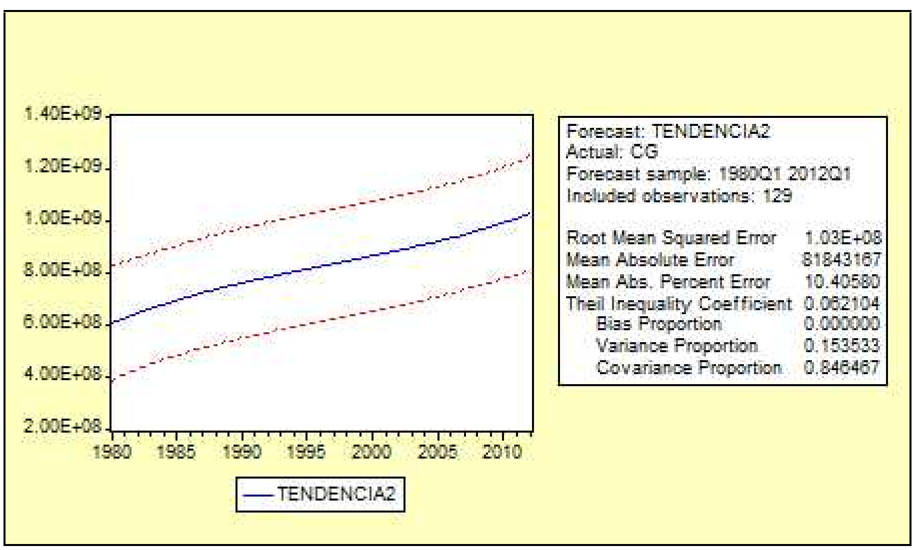
En virtud de que lo que se quiere en este análisis es conocer el comportamiento de la tendencia, por ello, ahora de la barra de herramientas de la regresión anterior se selecciona la opción Forecast, a continuación aparece un cuadro de diálogo, en él sólo se cambiará el nombre de la estimación: Forecast name: Tendencia2. A continuación se presenta el cuadro ilustrativo.

Cuadro 10.5.13 Descripción



De ese cuadro de diálogo se obtiene el siguiente recuadro que indica el comportamiento de la tendencia del CG.

Gráfico 10.5.7 Tendencia



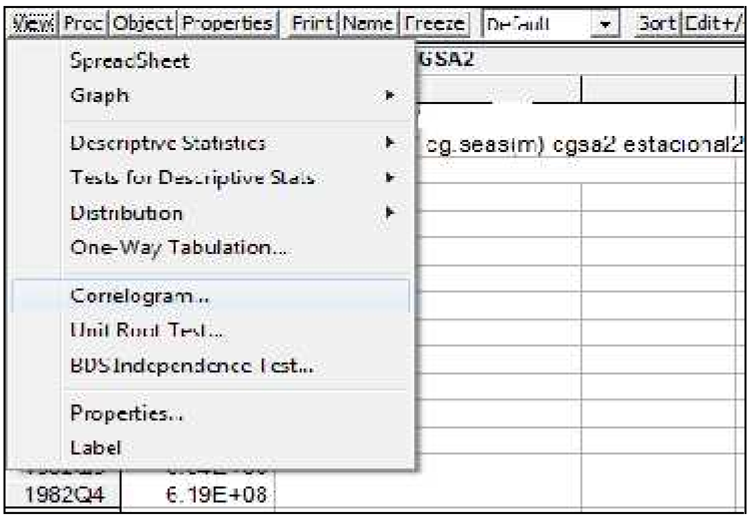
Entre otras cosas, se aprecia una tendencia ascendente y constante, por lo que se puede decir de manera reiterada que el Consumo Gubernamental ha crecido conforme aumenta el tiempo, tal y como se ilustró en el gráfico anterior.

Componente Ciclo.

Para obtener el componente ciclo es necesario generar la variable CGST2. En la barra de texto de Eviews 5 se escribe: genr cgst2=cgsa2-tendencia2

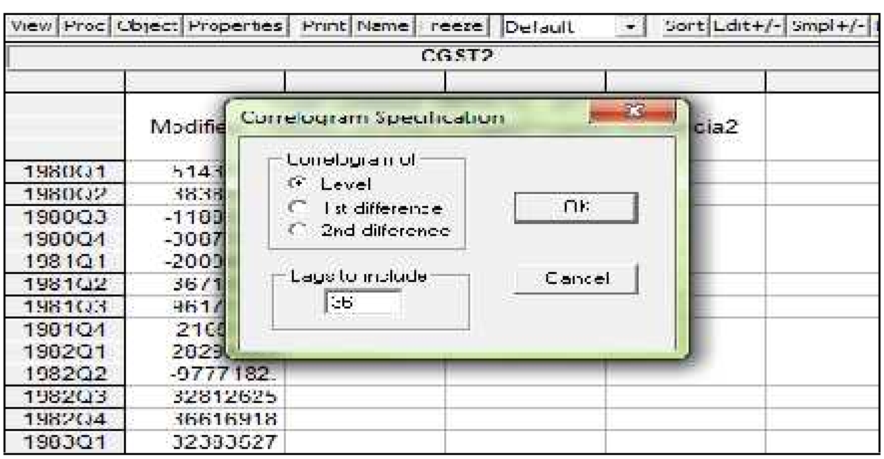
Una vez generada, es necesario abrirla y de ahí pasar a la opción: view/correlogram, tal como muestra a continuación.

Cuadro 10.5.14 Descripción



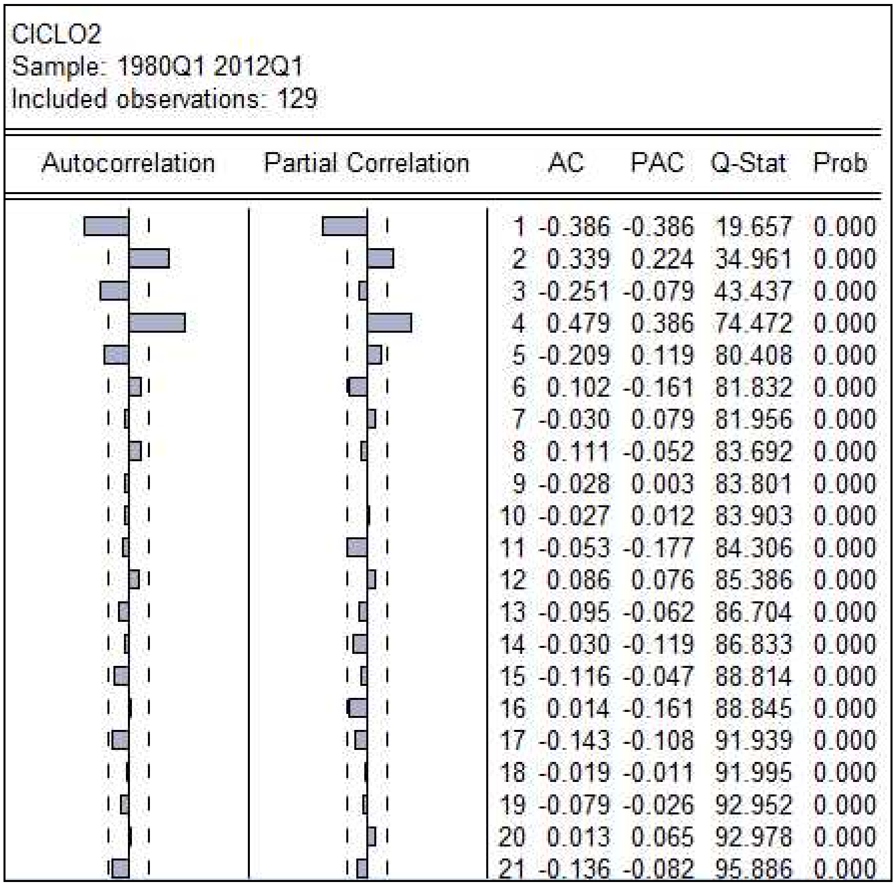
A continuación aparece un cuadro de diálogo, en la opción Correlogram of: seleccionar la opción level/ ok. Como se puede ver en el cuadro siguiente.

Cuadro 10.5.15 Descripción



Esta operación permitirá ver la correlación que tiene la variable cgst2, la cual dará pie al componente cíclico. Una vez realizados todas las operaciones para obtener el correlograma, se despliega una gráfica como la que se presenta a continuación.

Cuadro 10.5.16 Autocorrelación



CICLO2
Sample: 1980Q1 2012Q1
Included observations: 129

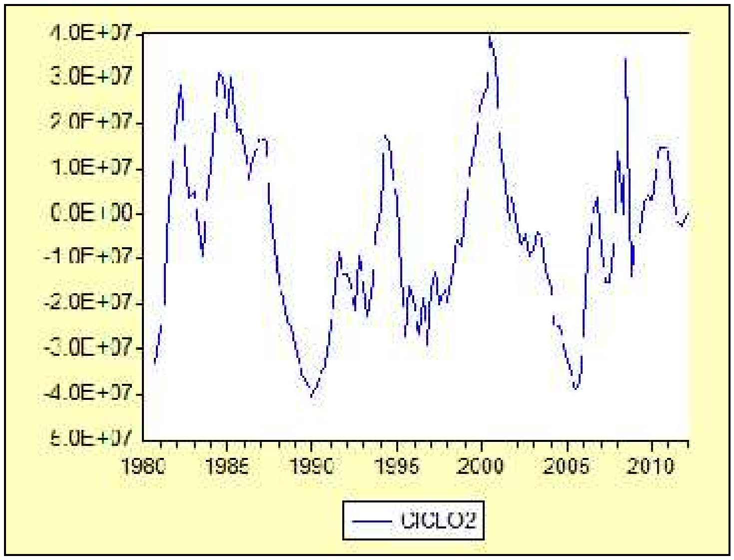
	AC	PAC	Q-Stat	Prob
1	-0.386	-0.386	19.657	0.000
2	0.339	0.224	34.961	0.000
3	-0.251	-0.079	43.437	0.000
4	0.479	0.386	74.472	0.000
5	-0.209	0.119	80.408	0.000
6	0.102	-0.161	81.832	0.000
7	-0.030	0.079	81.956	0.000
8	0.111	-0.052	83.692	0.000
9	-0.028	0.003	83.801	0.000
10	-0.027	0.012	83.903	0.000
11	-0.053	-0.177	84.306	0.000
12	0.086	0.076	85.386	0.000
13	-0.095	-0.062	86.704	0.000
14	-0.030	-0.119	86.833	0.000
15	-0.116	-0.047	88.814	0.000
16	0.014	-0.161	88.845	0.000
17	-0.143	-0.108	91.939	0.000
18	-0.019	-0.011	91.995	0.000
19	-0.079	-0.026	92.952	0.000
20	0.013	0.065	92.978	0.000
21	-0.136	-0.082	95.886	0.000

El Cuadro 10.5.16 muestra correlaciones simples y parciales, recordando que si alguna de las barras se sale de las bandas, entonces la serie presenta problemas de autocorrelación y, dependiendo del número de barras que se salgan de las bandas será el orden de autocorrelación. En la columna de correlación simple (Autocorrelation), el orden de correlación es cuatro. En la columna de correlación parcial, el orden de correlación podría ser de uno, ya que si se incluye un proceso autorregresivo de orden 1 AR(1) se corregiría. Otro método de detección de autocorrelación es la probabilidad asociada, la cual es significativa en todos los rezagos incluidos en este caso.

Para poder generar la variable ciclo se escribe en el cuadro de texto de Eviews 5:

genr ciclo2= @movav(cgst2, 4). Por si se pensaba que el paso anterior era inútil, nos sirvió observar el correlograma para saber cuántos rezagos incluir en la creación de la variable ciclo2. Una vez que se tiene la variable ciclo2, se abre, opción view/graph/line y se obtiene el siguiente cuadro.

Gráfico 10.5.8 Comportamiento de la serie

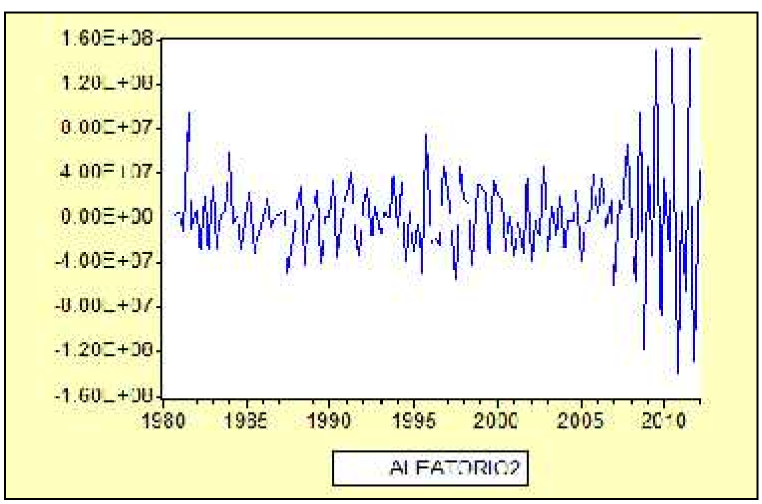


Como se puede ver el ciclo se presenta de manera irregular, pero tiene un comportamiento parecido al componente ciclo obtenido en la descomposición aditiva, lo cual constata que se ha trabajado bien.

Componente Aleatorio.

En el cuadro de texto de Eviews 5: genr aleatorio2=cgst2-ciclo2, se abre la serie, view/graph/line. El gráfico resultante se presenta a continuación.

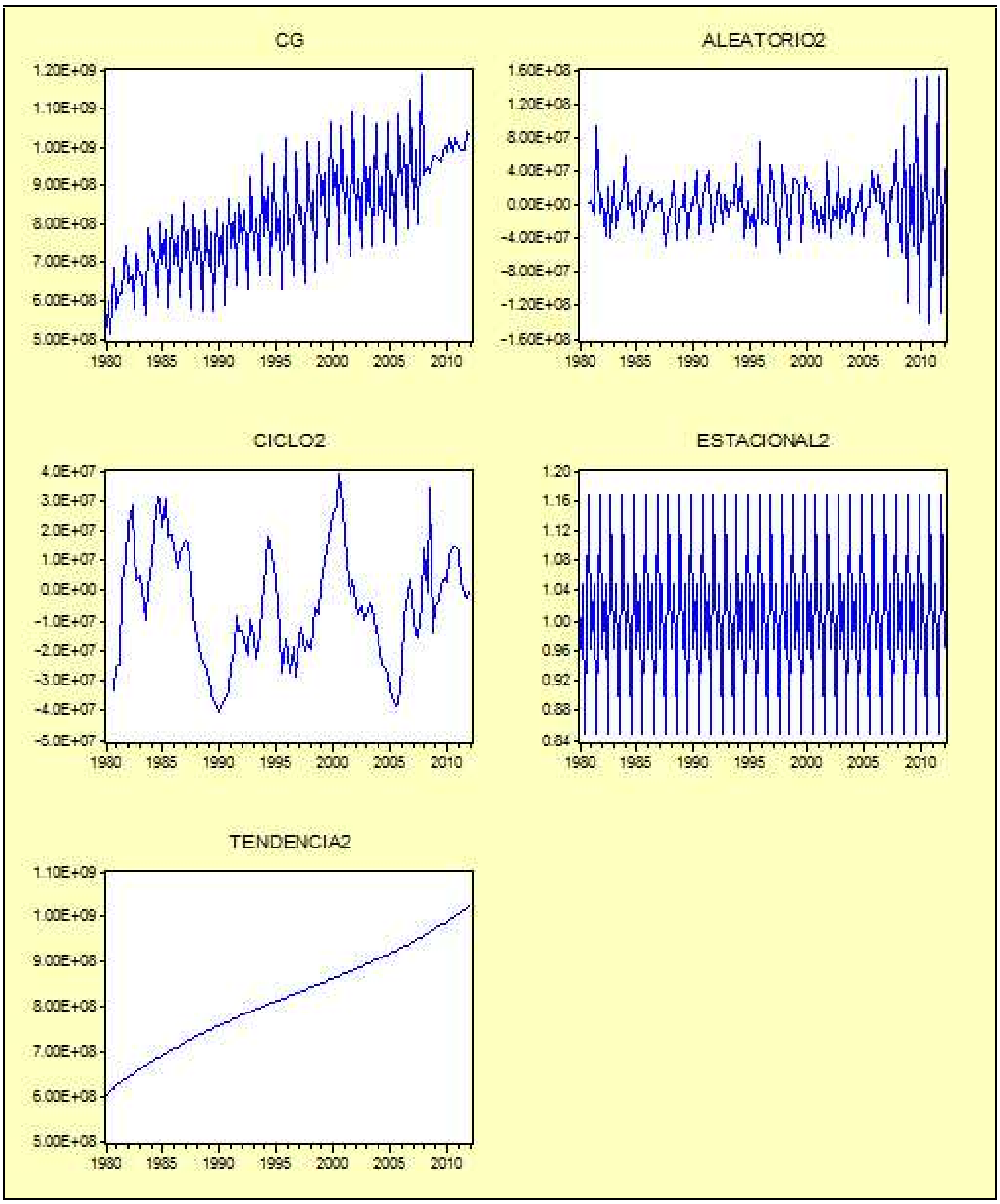
Gráfico 10.5.9 Comportamiento de la serie



Como ya se dijo, se espera que la esperanza matemática de los errores sean lo más cercanas a cero, sin embargo, este no es el caso, lo que nos indica que la serie tiene un componente irregular que se acentúa en 1981 y el periodo de la crisis actual 2009-2012.

Para obtener todas las gráficas de este método y, a manera de resumen, se seleccionan en el workfile cg, estacional2, tendencia2, ciclo2 y aleatorio2, con click derecho, open/as group, se abre la serie, en la opción view/multiple graph/line. El resultado de dicha operación es el siguiente:

Gráfico 10.5.10 Comportamiento de la serie



El fin de obtener estos gráfico es el de compararlos con los que se presentaron en el Método aditivo. En este caso, la tendencia es más clara que en el anterior, el ciclo se ve muy irregular, sin embargo, marca los mismos picos de auge que el método pasado. En cuanto los componentes estacionales y aleatorios, el comportamiento es exactamente igual.

CAPITULO XI.- ESTACIONARIEDAD Y RAICES UNITARIAS

Estos dos conceptos, como el anterior de estacionalidad, emanan del estudio que se hace del comportamiento de las series de tiempo. Su uso está fuertemente vinculado entre sí, por eso es que en el guión temático comprometido en el PAPIME 305811 para DGAPA, se presentan juntos. En este contexto es que a continuación empezamos con:

XI.1.-Estacionariedad

Es un concepto que es sinónimo de la estabilidad que muestran las fluctuaciones de una variable en un tiempo determinado por el investigador. Estadísticamente dicha estabilidad se expresa como las fluctuaciones que tienen los valores de la variable en torno a una media aritmética y sus dispersiones o alejamiento correspondiente, se observa que tienen una varianza y covarianza constantes. Cuando esto último se cumple, se dice que estas medidas estadísticas son invariantes en el tiempo, que la variable es estable o estacionaria estadísticamente en su tiempo de estudio.

En lo que atañe a su identificación, es decir, la identificación de si las variaciones de la variable son estables en el tiempo, empezaremos diciendo que los métodos que se utilizan son los que son objeto de estudio, caracterización y evaluación de su alcance en los siguientes incisos; baste decir ahora que son los gráficos, los numéricos como el correlograma, los modelos ARMA,ARIMA, las pruebas de Dicker- Fuller, Phillips-Perron, de raíz unitaria, etc.

Así con el fin no hacer cálculos equivocados ni interpretaciones erróneas de sus resultados, es recomendable que se estudie con detenimiento esta temática, ya que al hacer econometría empírica ésta plantea diversos desafíos a los investigadores cuando quieren hacer por ejemplo, econometría aplicada sobre análisis estructural, de predicción o de evaluación de políticas públicas.

No estudiar LA ESTACIONARIEDAD de una variable en el tiempo ocasiona lo siguiente:

Que las pruebas de hipótesis t, F, chi-cuadrada, etc., no proporcionen información adecuada para tomar decisiones;

Ignorar que la autocorrelación puede generarse simplemente porque las variables subyacentes no son estacionarias: correlación serial;

Aceptar la correlación espuria o falsa entre dos variables;

Aceptar la caminata aleatoria que en predicción, ergo, el precio de una acción es igual a su precio actual más un choque (innovación, crisis, etc.) puramente aleatorio, lo que hace elemental el método de pronosticar;

Los modelos de regresión que usan series de tiempo se utilizan para pronóstico, el cual será bueno si la serie temporal u variable es estacionaria. Al respecto se dice que es no estacionaria cuando su μ, σ2 y covarianza varían en el tiempo, por lo que decimos que no son estacionarias respecto a μ (primer momento) y con respecto a σ2 (segundo momento);

Que no se pueden hacer correctamente las pruebas DE CAUSALIDAD DE GRANGER Y SIMS; por consiguiente primero se deben hacer las pruebas de estacionariedad antes que las de causalidad.

¿Porqué corregir las series temporales y cómo? Roberto Montero Granados de la Universidad de Granada (2007) comenta que al medir la relaciones entre las variables surge el problema de que éstas observan una tendencia temporal, que de no resolverse puede generar "relaciones completamente espurias"; así, cuando las variables tienen una tendencia temporal definida se denominan " variables no estacionarias" o integradas. Para eliminar su relación espuria se cointegran. Luego cuando dos variables no estacionarias son cointegradas, sus residuos (diferencia o distancia entre ellas, como los llama Granger) producen una variable estacionaria, situación que hace que las estimaciones de variables no estacionarias sean consistentes, satisfactorias.

XI.1.1.-Pruebas para identificar la estacionariedad en las series temporales.

En este caso se explicará cómo se identifica la estacionariedad débil o covarianza:

Método gráfico.

Se grafica la serie temporal Yt, si su evolución muestra por ejemplo, una tendencia ascendente (o descendente), ello sugiere que su μ y σ2 están variando, indicando que quizás Yt es no estacionaria. Con este indicio se puede pasar al siguiente método gráfico más elaborado:

Correlograma y función de autocorrelación muestral (FAC).

Si definimos ρ como la autocorrelación entre los términos y k como diferencia entre ellos o rezago i-ésimo, entonces la autocorrelación ρ en el rezago k se determina así:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{Covarianzaalrezagok}{Varianza}$$

Cuando k =0, entonces ρ_0 = 1 ¿Porqué?, y que al estar la varianza y covarianza expresadas en las mismas unidades de medida, ρ_k al tener un significado como el del coeficiente de correlación, su valor oscila entre -1 y +1. Cuando se gráfica ρ_k versus k, la gráfica que resulta se denomina CORRELOGRAMA POBLACIONAL.

Recuerde que en la práctica no se trabaja la población sino una muestra suya, luego en este caso sus estimadores son:

$$\hat{\gamma}_k = \frac{\sum (Y_t - \overline{Y})(Y_{t+k} - \overline{Y})}{n}$$

$$\hat{\gamma}_0 = \frac{\sum (Y_t - \overline{Y})^2}{n}$$

Donde n es el tamaño de la muestra y $\overline{Y}$ es la media aritmética de la muestra. Así, la función de autocorrelación muestral al rezago k es:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$$

La gráfica de $\hat{\rho}_k$ en el eje vertical y k en el eje horizontal del primer cuadrante del sistema de ejes cartesianos ahora se llama CORRELOGRAMA MUESTRAL, que construida con EViews generalmente muestra la AC (función de autocorrelación muestral) y ACP (función de autocorrelación parcial), más la estadística Q y su probabilidad de ocurrencia asociada. AC tiene una línea vertical continua que representa el eje cero a partir del cual se grafican los diferentes valores positivos (a la derecha) y negativos (a la izquierda) de ρ_k: Coeficiente de autocorrelación individual.

Cuando el proceso estocástico es puramente ruido blanco es estocástico en distintos rezagos que se localizan cerca del cero, lo que significa que se trata de una serie temporal estacionaria. Al respecto, cuando el proceso o serie temporal es MCA las ρ_k se ubican cerca de uno indicando que se trata de una serie no estacionaria.

Cabe señalar que el número de rezagos a exhibir es discrecional; lo importante es que dé información suficiente para identificar la estacionariedad de la serie. En opinión de Gujarati (2003: 784) la elección del número de rezagos es una decisión empírica ya que en la vida real se sugiere calcular las FAC en un número próximo a un tercio o una cuarta parte del número total de elementos que integran la serie temporal. No obstante, autores como Carrascal et al sugieren aplicar un criterio estadístico para determinar su número; pueden ser el criterio de información Akaike o el de Schwarz. <u>Como opciones también se sugieren prueba como las siguientes:</u>

2.1. Verificar la significación estadística de los coeficientes de autocorrelación individuales, ρ_k´s. Y establecemos:

Ho: $\rho_k = 0$; no hay autocorrelación entre los términos cuando la probabilidad es > 5% indicando que la serie es estacionaria;

Ha: $\rho_k \neq 0$; si hay autocorrelación entre los términos cuando la probabilidad es < 5% indicando que la serie es no estacionaria.

Cada uno de los ρ_k se prueba o contrasta. Al respecto, Bartlett demostró que si una serie de tiempo es puramente aleatoria, es decir, si muestra ruido blanco las ρ_k´s muéstrales son aproximadamente:

$\hat{\circ}_k \sim N(0, 1/n)$

Lo que indica que se distribuyen normalmente y que sus parámetros son $\mu = 0$ y σ^2 = 1/tamaño de la muestra (Gujarati, 2004: 788). Luego, cuando por ejemplo, n =88, la varianza igual a 1/88 = 0.01136 y su $\sigma = \sqrt{0.01136} = 0.1066$. Con ello construimos el intervalo de confianza con α = 5% para cualquier ρ_k, así:

$$\hat{\circ}_k \pm Z_{\sim} \pm \circ_k$$

Si ahora hacemos k = 10 y sí su ACF = 0.638 = $\hat{\circ}_{10}$

Sustituyendo: $\hat{\rho}_{10} \pm 1.96\ (0.1066)$

$\hat{\rho}_{10} \pm 0.2089$

Así, 0.638 ± 0.2089

INTERPRETACIÓN: El intervalo de confianza del 95% para la verdadera ρ_{10} va de 0.4091 a 0.8469. Al respecto, *cuando el intervalo de confianza no incluye cero se rechaza la Ho de que la verdadera ρ_k es cero*. Como en este ejemplo numérico el intervalo de confianza no incluye cero, ello indica que debemos rechazar Ho de que al verdadera $\rho_k = 0$; en otras palabras, se tiene 95% de confianza de que la verdadera ρ_{10} sea significativamente distinta de cero. Concluimos diciendo que la serie temporal es no estacionaria.

2.2. Probar la significación estadística del ***conjunto*** de ρ_k´s en lugar de que cada ρ_k individual. Lo anterior implica suponer que todas las ρ_k son simultáneamente iguales a cero, lo cual se verifica con la **estadística Q**, que desarrollaron Box & Pierce y definieron como:

$$Q = n\sum_{k=1}^{m} \hat{\rho}_k^{2}$$

Donde m = longitud del rezago

n = tamaño de la muestra

k = rezago i-esimo

Aquí se establece:

Ho: $\rho_1 = \rho_2 = \ldots \rho_k = 0$; la serie es estacionaria porque no hay autocorrelación conjunta.

Ha: $\rho_1 \neq \rho_2 \neq \ldots \rho_k \neq 0$; la serie es no estacionaria porque sí existe autocorrelación conjunta.

Así, con un valor de α determinado, la estadística Q suele por lo general aplicarse cuando se supone que la serie temporal es ruido blanco, de la cual pueden extraerse grandes muestras, como sucede con la ji-cuadrada.

Para tomar decisiones sobre si aceptamos o rechazamos la hipótesis nula se acostumbra comparar la Q calculada o real con la Q_α teórica o de tablas. Si $Q > Q_\alpha$ de la tabla de ji-cuadrada con α definida en determinado nivel de significación, se rechaza Ho de que todas las ρ_k son iguales a cero.

Como alguna Q puede ser diferente de cero en la vida real y ello introduce la duda de qué hacer en ese caso, Ljung-Box elaboraron como una derivación de la Q de Box & Pierce, **la estadística LB** por las iniciales de sus autores. Así, EViews calcula además del correlograma y las ρ_k´s a la estadística Q, con su probabilidad de ocurrencia asociada (descrita en la última columna). Luego, si establecemos:

Ho: $\rho_1 = \rho_2 = \ldots \rho_k = 0$

Ha: $\rho_1 \neq \rho_2 \neq \ldots \rho_k \neq 0$

Con α = 5%, decimos que si la probabilidad de Q (LB) es menos o igual al 5%, aceptamos Ha y decimos que la serie es no estacionaria, y si es mayor que 5% aceptamos Ho y decimos que la serie es estacionaria, que es ruido blanco y que sus μ y σ^2 son constantes en el tiempo.

La formula de Q de Ljung-Box viene dada por:

$$LB = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{P}_k}{n-k}\right) - X^2 m$$

Cabe señalar que en muestras grandes Q y LB tienen una distribución Chi cuadrada con m grados de libertad; sin embargo se ha observado que LB en muestras pequeñas tiene más potencia estadísticamente hablando. Al usar Eviews para calcularla, en el cuadro además de aparecer la FAC y FACP de las "ro" y sus correspondientes correlogramas, también aparece la estadística Q ahora llamada LB en la penúltima columna con su probabilidad de ocurrencia (última columna). Al respecto, cuando dicha probabilidad es *cercana a cero, se dice que la probabilidad de que LB tome el valor de Q en determinado rezago bajo la hipótesis nula de que la suma de los k cuadrados (columna de FAC) sea cero,* ***es*** muy pequeña (última columna) e indica también que la serie temporal es no estacionaria, ya que sabemos que para que sea estacionaria su probabilidad debe ser mayor a 5%.

2.3.- FACP: función de autocorrelación parcial.

Por otra parte, el análisis de series temporales además de AC, se complementa con ACP: autocorrelación parcial. Por su singularidad a continuación transcribimos el algoritmo que Pérez (2007: 550) sugiere para su cálculo:" El primer término de la función de autocorrelación parcial, que vamos a denotar por θ_{11}, puede estimarse transformando la serie X_t en desviaciones respecto a su media muestral ($Y_t = X_t$ menos su media) y a continuación estimando una regresión de Y_t sobre Y_{t-1}. La pendiente estimada de esta regresión es θ_{11}. El modelo de regresión es $Y_t = \theta_{11}Y_{t-1} + U_t$. Además, el primer valor de la función de autocorrelación parcial θ_{11} es precisamente igual al primer valor de la función de autocotrrelación. Esta es una propiedad de las funciones de autocorrelación de todo proceso estocástico estacionario.

El segundo valor de la función de autocorrelación parcial, θ_{22}, se estima mediante una regresión de Y_t sobreY_{t-1} eY_{t-2}. El modelo de regresión es $Y_t = \theta_{21}Y_{t-1} + \theta_{22}Y_{t-2} + U_t$.

El tercer valor de la autocorrelación parcial, θ_{33}, se estima mediante una regresión de Y_t sobre Y_{t-1}, Y_{t-2} e Y_{t-3} . El modelo de regresión es $Y_t = \theta_{31}Y_{t-1} + \theta_{32}Y_{t-2} + \theta_{33}Y_{t-3} + U_{t3}$.

Vemos pues que la función de autocorrelación parcial puede estimarse mediante una series de regresiones, cada una de las cuales contiene como variable explicativa un retardo más que la anterior, y de que nos vamos quedando en cada caso con los coeficientes estimados en los retardos más altos: θ_{11},θ_{22}, θ_{33} …….., que son así los valores estimados de la función de autocorrelación parcial estimada mediante fórmulas recursivas, utilizando la función de autocorrelación previamente estimada y utilizando las ecuaciones de Yule-Walker. A veces se suele denominar c*orrelograma* a la representación gráfica de las funciones de autocorrelación parcial."

Bien, después de esta excelente exposición de Pérez, ahora abordemos otro método para verificar si la serie temporal es o no es estacionaria.

XI.2.- Prueba para identificar la raíz unitaria en las series temporales.

Recordando el planteamiento y definición dados en párrafos anteriores, ahora profundizaremos para arribar a una expresión algebraica que nos permita visualizar mejor la raíz unitaria, su comprensión, interpretación y corroboración usando otros signos, de los que antes se dijo; claro ésta vista ahora en el contexto de la prueba de hipótesis que indicará si la serie temporal es o no estacionaria.

Así, recordando que si establecemos que $\mathbf{Y_t = \rho Y_{t-1} + U_t}$ donde U_t es el término de error con ruido blanco y que ρ toma valores entre -1 y +1. Al respecto, se dijo que cuando ρ = 1 la ecuación anterior se convierte en MCA sin variaciones porque es un proceso estocástico no estacionario (Gujarati ,2004: 788), i.e., la serie es no estacionaria. Para eliminarla se puede hacer la regresión de Y_t en función exclusivamente de su primera diferencia o valor rezagado anterior (Y_{t-1}) y con ello observar qué valor toma ρ, tal que, por ejemplo, si es cero entonces Y_t es estacionaria con ruido blanco; si fuera ρ = 1 entones Y_t es no estacionaria. Con ese fin Gujarati (idem) hace las siguientes manipulaciones algebraicas:

$$Y_t - Y_{t-1} = {}^\circ Y_{t-1} - Y_{t-1} + U_t$$

$$Y_t - Y_{t-1} = ({}^\circ - 1)Y_{t-1} + U_t$$

Que ahora expresamos como:

$$\Delta Y_t = \text{¢} Y_{t-1} + U_t$$

Donde δ es el coeficiente estimado de la pendiente de Y_{t-1} la cual actúa como variable regresora para estimar Y_t, que a su vez procuramos que sea estacionaria. Sabemos que Δ es el operador de la primera diferencia y que si δ = (ρ – 1) entonces se puede probar:

Ho: δ = 0; Y_t es no estacionaria, porque ρ=1 y se habla de un MCA: proceso estocástico no estacionario.

Ha: δ ≠ 0; Y_t es estacionaria, con α = 5%.

Luego cuando δ = 0, entonces ρ = 1, **<u>se tiene una raíz unitaria</u>** y se concluye con que Y_t es no estacionaria; en este sentido recuérdese que cuando δ = 0

$$\Delta Y_t = (Y_t - Y_{t-1}) = U_t$$

Al expresar U_t la autocorrelación de uno con otro término y al ser a la vez U_t un término de error con ruido blanco, se dice que no hay autocorrelación entre los dos términos de Y, por lo que al igual que en párrafos anteriores, se concluye con que "las primeras diferencias de una serie de tiempo de caminata aleatoria, son estacionarias".

Por otra parte, obsérvese que cuando δ es negativo entonces Y_t es estacionaria porque si δ = ρ – 1, para que δ sea negativa ρ debe ser menor que uno, ergo, ρ = 0.10, así δ = 0.10 – 1 = -0.9, luego Y_t es estacionaria.

XI.2.1- ¿Cómo saber si δ es o no cero? La respuesta la **dieron Dickey-Fuller (DF) con** su estadística Tau (^{2}), cuyo uso se ilustrará más adelante.

a).-

XI.2.1.1.- Pruebas para verificar si la serie temporal es o no estacionaria.

a) Visual, identificar si los gráficos de las variables muestran que éstas crecen o decrecen monótonamente (cuando los shocks son persistentes) o por el contrario, no es posible identificar un patrón de comportamiento definido. Entonces pues es importante verificar si la serie presenta <u>componente tendencial;</u>sí éste se presenta la variación de los términos de la serie a lo largo del tiempo <u>no es en torno a un valor constante</u>(Peña, 2005:18) por lo que <u>decimos que no es estacionaria **"en media"**</u> o, <u>si la dispersión de</u>

los datos no es parecida en el tiempo, se dice que la serie es no estacionaria en **varianza.** Al respecto, para verificar la no constancia de la varianza en el tiempo se usa la *gráfica llamada Rango- Media.*

Así, cuando hay *estacionariedad en la media aritmética, usando el operador diferencias la gráfica de la serie debe tender a una línea recta paralela al eje de las "x"; cuando no hay estacionariedad se observa cierta tendencia(ascendente o descendente) que evidencia que la varianza de la serie de datos no es constante, i.e., que no hay homogeneidad en las variaciones de sus términos a lo largo del tiempo.*

b).-Solución a la falta de estacionariedad en la serie temporal:

Para eliminar la media y la varianza "variables", los datos deben transformarse en logaritmos y luego hacer en estos nuevos datos el *número de diferencias necesarias para estacionarizar* la serie temporal. De acuerdo a Pulido y López (1999: 264) en *términos gráficos*, una **media constante** supone la no existencia de tendencia (una línea de tendencia paralela al eje de las abscisas) y una **varianza constante** corresponde a un gráfico en que las oscilaciones alrededor de la media *sean similares*, lo que técnicamente se conoce como homoscedasticidad, i.e., las observaciones tienen una misma varianza a lo largo del tiempo.

Para conseguir una serie estacionaria **en media** es necesario efectuar *diferencias sucesivas* de la misma, generalmente *es con la primera diferencia* en tanto que es *con la segunda* para hacerla estacionaria en **varianza**. Por su parte, una manera sencilla de reducir la heteroscedasticidad, es decir, transformar una serie en otra con varianza relativamente constante, *es tomar logaritmos,* ya que ello supone trabajar con valores relativamente más homogéneos. Estas transformaciones logarítmicas de datos son muy frecuentes en los modelos ARIMA (metodología de Box & Jenkins), principalmente unidas a una diferencia posterior de la serie transformada, es decir, en logaritmos.

Luego entonces, corroborando lo antes dicho, por lo que respecta a transformar una serie estacionaria en **media**, una tendencia lineal será corregida tomando la primera diferencia; mientras que una tendencia no lineal **(varianza**) suele llevar en la práctica al uso de dos o más diferencias. Por ejemplo, si una serie muestra una tendencia lineal, su primera diferencia ya no tendrá esa tendencia. La eliminación de una tendencia cuadrática puede conseguirse mediante doble diferenciación (Pérez, 2003: 269)

Asimismo, también se aconseja trabajar con el *correlograma* puesto que también tiene una sección gráfica a de la función de autorregresión (ρ_k), donde, reiterando lo dicho párrafos arriba:

ρ_k = Cov (X_t ,$X_{t+k)/}$| [var (X_t) var(X_{t+k})]$^{1/2}$; v_k = 1, …, m

Así, a manera de corolario en lo que se refiere al corrrelograma, decimos que cuando los valores de los coeficientes de autocorrelación de AC y PAC del correlograma descienden lentamente, ello indica variables no estacionarias. Cuando dichos valores descienden rápidamente de manera que se acercan a cero ello indica que ha disminuido la autocorrelación, se dice que es una situación cuasialeatoria y que corresponde a variables estacionarias.

Cabe mencionar que estas pruebas visuales son calificadas como no formales y por consiguiente que no son definitivas. Por eso a continuación se exponen las siguientes:

b) Prueba de Dickey-Fuller, (DF)

Tiene mayor rigor estadístico que las pruebas anteriores dado que su hipótesis nula no consiste en indagar si la serie es "ruido blanco", sino que pretende detectar si la serie tiene "raíz unitaria"; en otras palabras, busca verificar si no existe relación entre el aumento de cada valor y el inmediato anterior, tal que si no hay relación la serie es estacionaria (I(0)); cuando si existe relación, se dice que la serie tiene raíz unitaria (I(1)).

Al respecto cuando se buscan relaciones lineales entre las variables, dichas relaciones lineales pueden ser con origen, sin origen, con tendencia, etc. Así, la prueba puede realizarse en tres formas distintas:

1. $dx_t = rX_{t-1} + U_t$

2. $dx_t = a_0 + rX_{t-1} + U_t$

3. $dx_t = a_0 + rX_{t-1} + a_{1t} + U_t$

Donde $dx_t = X_t - X_{t-1}$; en los tres casos se establece H_0: $\rho = 0$ y Ha: $\rho \neq 0$

Si se acepta Ho (p-valor $\geq$ 0.05) la serie es estacionaria y tiene raíz 0 (I(0)); si se rechaza Ho (p-valor $\leq$ 0.05) la serie es no estacionaria y tiene raíz unitaria 1(I(1)).

Donde **p** es la probabilidad de que el parámetro sea significativo estadísticamente. Cuando se calcula con Eviews generalmente aparece en la última columna del cuadro que muestra éste y otros datos más.

Es importante comentar que esta prueba es correcta sólo en los casos que la correlación de la variable a "testar" no contiene *retardos* en la correlación. Para mejorarla los autores formularon la "prueba aumentada" **(DFA**). Su diferencia con la anterior es que en cada regresión se incluyen retardos de la variable como independientes.

¿Pero cuál es el número apropiado de retardos (k)?

En opinión de Verbeek, cuando es muy pequeño o muy grande puede hacer que series no estacionarias aparezcan como estacionarias. Pueden usarse: **a**) El Criterio de la Información de Akaike (AIC) o **b**) Criterio de Schwarz Bayesian (SBC); en los dos casos se escoge la de menor valor (Gujarati, 2004) y con mayor Coeficiente de determinación (Carrascal et al, 1999); **c)** Incluir k´s mediante prueba y error hasta eliminar la CORRELACION SERIAL en las Ut´s que, como se recordará, se mide con el estadístico Durbin –Watson; **d**) Usar la B de Bartlett, Q de Parmateau, también conocida como prueba Ljung-Box y la Z de Phillips- Perron.

En estos casos las pruebas de hipótesis se establecen así:

Ho: (p-valor≥0.05) la serie es "ruido blanco"= estacionaria, i.e. $\rho=0$; y

Ha: (p-valor≤0.05) la serie es "caminata aleatoria" = no estacionaria, $\rho\neq0$.

En que **p** es la probabilidad de que el parámetro en estudio sea estadísticamente significativo, que cuando se usa EVIEWS, el cuadro que produce con diferentes conceptos y sus valores, generalmente el valor de "p" aparece en la última columna del cuadro con valores específicos para su probabilidad.

c).- Prueba de Phillips- Perron

En el caso de Phillips- Perron (1988) se calculan parámetros Zp y ZŁ, donde Ł: *es la estadística tau, c*uyos valores críticos y significación correspondiente se basan en la misma t modificada de la prueba DF. Sus ventajas sobre las pruebas anteriores son:

1.- La Ho consiste en que la serie es integrada de orden 1, i.e., con una p cuyo valor ≤ 0.05 se rechaza la hipótesis nula, lo cual indica que la serie no es I(1).

2.- Se pueden introducir rezagos (lags en inglés) en la prueba, situación que lo hace insustituible en el caso de que no haya sido recomendada la introducción de los mismos.

d).-Comentarios:

Es importante señalar que para Gujarati (2004) los **siguientes conceptos**: serie estacionaria, ruido blanco y raíz nula (I(0)), por una parte, así como por otra parte, serie no estacionaria, caminata aleatoria y raíz unitaria (I(1)), respectivamente, **por lo general son sinónimos entre sí;** pero en opinión de Montero (2007) **no son exactamente sinónimos entre sí: "**pero si aluden a distintas formas del comportamiento de las variables (ciclos, estacionalidad, tendencias, etc.) que pueden hacer que las variables puedan ser útiles o completamente inútiles". En este contexto también debe decirse que en opinión de Perron (1989) las pruebas antes descritas son ineficientes cuando sucede un cambio estructural (¿situación actual?), i.e., una series estacionaria con cambio estructural puede mostrarse como no estacionaria y viceversa.

Al respecto, por la importancia de la didáctica usada en la enseñanza-aprendizaje, para enfatizar las diferencias entre el análisis de serie univariables y multivariables, **debe reiterarse q**ue dos o más variables están cointegradas cuando cumplen determinados requisitos como el que la perturbación resultante de su relación lineal, Ut, sea estacionaria. Para que esto suceda es necesario que las variables integradas (no estacionarias) sean del mismo orden y que adicionalmente, estén cointegradas, i.e., que estén vinculadas mediante una relación lineal estable tal que Ut, diferencia entre el valor real de la variable dependiente y su valor estimado, sea estacionaria. O sea que hacer análisis de cointegración entre dos o más variables equivale a analizar la estacionariedad de la perturbación aleatoria, Ut, del modelo que las relaciona. Este análisis de estacionariedad de una variable, Ut, se realiza con al ADF: Prueba aumentada de Dickey-Fuller.

XI.3.-Profundización conceptual

En virtud de que estamos introduciendo nuevos conceptos que en mucho emanan del estudio de la estacionariedad de las series temporales, es conveniente profundizar profundizando en los ya mencionados y/o introduciendo otros, como los siguientes:

Modelos Dinámicos o Autorregresivos.

Se llaman dinámicos porque no son estáticos, es decir, sus valores cambian, están relacionados con la variable *tiempo*, evolucionan conforme pasa el *tiempo*.Los hay de dos tipos:

a).- Se denominan autorregresivos cuando el modelo incluye uno o más valores pasados (rezagos) de la variable dependiente entre sus variables independientes: ejemplo: $Y_t= \alpha+\beta X_t + ɠY_{t-1} + Ui$ ……………..(1)

b).- Se llaman modelos de rezagos distribuidos cuando el modelo de regresión incluye sólo los valores actuales y pasados (rezagados) de las variables explicativas en el lado derecho de la ecuación, ergo: $Y_t=\alpha+\beta_0 X_t +\beta_1 X_{t-1} + \beta X_{t-2} + Ui$ ……………(2)

Se habla de rezagos porque una variable regresada (Y) puede responder a una o varias regresoras (X_1, X_2, X_3, …., X_n) *con un rezago de tiempo.*

En este contexto se dice que para estimar los modelos de KOYCH y deExpectativas Adoptativas Consistentemente, se usa el método de *variables instrumentalesque ya se* vio en análisis de autocorrelación. Como se recordará, la variable instrumental es una variable representante para la variable regresada que tiene la propiedad de que no está correlacionada con el término error: Ui.

III.3.1. Modelos con variables retardadas.

Se dice que la incorporación de efectos diferidos en el tiempo en un modelo econométrico se hace a través de variables retardadas, que se manejan en dos tipos de modelos que se manipulan con tratamientos econométricos diferentes: los modelos autorregresivos y los de retardos distribuidos. Ver a) y b) anteriores.

III.3.1.1. Modelos autorregresivos.

Son aquellos en que el efecto se recoge mediante la incorporación de la variable dependiente retardada en algún periodo, como variable explicativa. Estos modelos suponen una violación de los supuestos básicos: la hipótesis o supuesto de que las variables explicativas no son aleatorias, en virtud de que la variable endógena retardada depende de la perturbación aleatoria y por consiguiente, tiene un carácter estocástico.

Al respecto debe comentarse que los resultados de la estimación MCO de un modelo con regresores estocásticos están condicionados al tipo de DEPENDENCIA que existe entre estas y las perturbaciones. La dependencia puede ser:

- Independencia total. En este caso se conservan todas las propiedades de los estimadores por MCO.
- Dependencia parcial, Sucede cuando el regresor estocástico sólo depende de la perturbación en periodos de tiempo pasados pero no en el presente ni en el futuro. En esta situación los estimadores son sesgados pero mantienen la propiedad de consistencia.
- Dependencia total. Aquí el regresor estocástico depende de las perturbaciones en todos los periodos. El estimador MCO, además de ser sesgado, ya no es consistente, por lo que se recurre a otro método de estimación alternativo que al menos garantice esta última propiedad asintótica. El método que utilizaremos es el de variables instrumentales (Carrascal, et al, 2000:294).

Con este método se sustituyen los regresores exogénos por ellos mismos así como la variable endogéna retardada por otra variable exógena con la que presente mayor correlación, retardada en el mismo número de periodos.

La determinación del tipo de dependencia entre el regresor estocástico (la variable endógena retardada) y la perturbación se realiza analizando la existencia de autocorrelación en el modelo. Es interesante señalar que la ausencia de correlación en las perturbaciones implica una situación de dependencia parcial, en tanto que cuando hay autocorrelación ello implica que hay dependencia total. Para detectar la autocorrelación se usa el contraste h de Durbin, que plante las mismas hipótesis que la DW de Durbin Watson:

H_o: No hay autocorrelación; =0

H_a: Hay autocorrelación, AR(1) positiva o negativa; <0 ó >0

$$h = \hat{\rho}\sqrt{\frac{T}{1-TS^2_{\hat{\beta}_{k-1}}}} \longrightarrow N(0,1)$$

Donde $\hat{\rho}$ estimador correspondiente a la regresión de los residuos MCO frente a sí mismos retardados un periodo y sin término constante. $S^{2}_{\hat{}\,k-1}$ es el estimador de la varianza del estimador correspondiente a la variable endógena retardada un periodo.

Ejemplo: con =5% si h>1.645 se rechaza Ho: y decimos que hay autocorrelación positiva. Cuando h<-1.645 se rechaza Ho y se comenta que hay autocorrelación negativa.

Como se indicó antes, cuando sucede lo anterior se usa el método de variables instrumentales, que implica buscar unos “instrumentos” que sustituyan a los regresores del modelo. Así en el caso de la variable endógena retardada se usa como instrumento uno de los regresores también retardado.

Para detectar la autocorrelación también se usa el correlograma de los residuos y el contraste de Breusch-Godfrey. Al eliminarse la autocorrelación con los métodos sugeridos se conoce el efecto en el corto como en el largo plazo de la variable endógena retardada. Ver Carrascal pp. 305 último párrafo.

XI.3.1.1.1.-Conceptos básicos:

Distribución de frecuencias: es la colocación o repartimiento de los datos de un fenómeno económico de acuerdo con el número de veces que se repite cada uno de ellos, como por ejemplo el peso o la estatura de las personas.

Distribución de probabilidad, es la lista de los resultados que se producen en la realización de un experimento, mismos que aparecen con su probabilidad de ocurrencia asociada a cada uno de ellos. Ejemplo, el experimento puede ser el lanzamiento de una moneda al aire o de un dado un determinado número de veces, cuyos resultados se registran junto con la probabilidad de ocurrencia asociada a cada uno de ellos.

Estacionalidad: es el efecto periódico que tiene una estación del año en los valores de una serie temporal, ejemplo: en México las compras en navidad son mayores que en el resto del año porque las personas reciben ingresos extras a los que normalmente recibe en el mes ordinario

Estacionariedad: Es lograr que una serie temporal registre media y varianza constantes durante el periodo de estudio de la misma.

Vector: Es un agente que transporta algo de un lugar a otro. En matemáticas es un segmento de recta orientado en que se distingue un origen y un extremo (final).

Vector fijo: Es un segmento cuya posición en el espacio está totalmente determinada.

Vector libre: Es un segmento que puede ser reemplazado por cualquier otro vector paralelo, del mismo sentido y magnitud.

Radio vector: Es un segmento orientado que tiene como origen un punto fijo, y cuyo extremo puede desplazarse sobre una curva dada.

Espacio vectorial: Estructura de un conjunto cuyo modelo lo proporcionan las propiedades de los vectores libres del espacio euclidiano ordinario.

Sistema de vectores deslizantes: Conjunto compuesto de un número finito de vectores móviles sobre su línea de aplicación.

Cálculo vectorial: Es el estudio de las funciones de una variable vectorial.

Variable vectorial: Cuando es de orden N es un vector compuesto de N elementos, cada uno de los cuales es una variable de un problema matemático (Web, 2009).

Magnitud vectorial: Magnitud física cuya definición exige un valor numérico, una dirección y un sentido.

Definiciones de **Escalar**:

- Se denomina escalar a los números reales, complejos o racionales que sirven para describir un fenómeno físico con magnitud, pero sin las características vectoriales de dirección o sentido. Formalmente es un tensor de rango cero.

En informática, un escalar es valor, variable o campo que sólo puede tener un valor en un cierto momento; en comparación, están los conceptos de array, lista u Objeto, que pueden tener almacenado en su estructura más de un valor.

Ejemplo: Una magnitud física se denomina escalar cuando puede representarse con un único número (única coordenada) invariable en cualquier sistema de referencia. Así la masa de un cuerpo es un escalar, pues basta un número para representarla (75 kg).

Variable determinista: Es aquella cuyo valor se conoce con certeza.

Variable estocástica o aleatoria: Es aquella que no se conoce su valor pero que se puede predecir con probabilidades dado que tiene un comportamiento parecido al de las distribuciones probabilisticas teóricas conocidas, cuyas características (o propiedades estándar) sirven para explicar el comportamiento de variables económicas; así, éstas últimas pueden describirse usando las probabilidades.

Variable aleatoria discreta, es aquella que sólo puede tomar ciertos valores que son indivisibles o no fraccionables.

Variable aleatoria continua, es aquella que toma cualquier valor en un intervalo determinado, es decir, toma valores fraccionables o divisibles.

Variable dummy, es una variable dicotómica porque toma sólo dos valores: 0 y 1; generalmente describe una cualidad particular (sexo, religión, etc.) y por consiguiente cuantifica dicha cualidad con objeto de analizarla como sucede con las variables cuantitativas.

Covarianza, es una medida estadística de la relación entre dos variables aleatorias:

Ella indica la forma en que las variables se mueven juntas; cuando es positiva indica que las variables se mueven en la misma dirección; cuando es negativa indica que se mueven en sentido contrario, que hay una relación inversa entre ellas y cuando es cero, indica que las variables no están relacionadas, que son independientes. La medición de la relación entre las dos variables es muy clara cuando se estandarizan sus datos porque entonces se *hace equivalente al coeficiente de correlación* que varía entre -1 y +1, de manera que cuando es igual a 1 indica que las dos variables varían en el mismo sentido; cuando es menos 1, indica que varían en sentido inverso (mientras una aumenta, la otra disminuye) y cuando es cero, indica que las variables no están correlacionadas.

XI.3.1.1.2.-Procesos estocásticos integrados.

Comenta Gujarati (2004) que MCA: Modelo de Caminata Aleatoria es un caso específico de una clase más general de procesos estocásticos conocidos como *procesos integrados*. Así, recordando que la MCA sin variaciones es no estacionaria, pero que su primera diferencia la hace estacionaria, se dice que MCA sin variaciones es un *proceso estocástico de orden 1* y se expresa como: I (1). Luego si una variable tiene que diferenciarse dos veces para hacerse estacionaria se toma la diferencia de la primera diferencia; a esta serie se le llama *serie temporal integrada de orden dos*, I (2), etc.

En este contexto, si una serie de tiempo no estacionaria tiene que diferenciarse “d” para hacerla ruido blanco o estacionaria, decimos que es una *serie integrada de orden” d”*; si ella se denota como Y_t entonces se escribe $Y_t \sim I(d)$.

Cuando Y_t desde el principio es estacionaria de manera que no es necesario diferenciarla, se le llama integrada de orden cero y se escribe $Y_t \sim I(0)$; se intuye entonces “que serie de tiempo estacionaria” y “serie de tiempo integrada de orden cero” significan lo mismo, i. e., son sinónimos.

XI.3.2-Propiedades de las series o variables integradas.

Si tenemos tres variables o series de tiempo Y_t, X_t y Z_t, entonces:

1. Una combinación lineal de series temporales estacionarias y no estacionarias, es estacionaria, ergo, si $X_t \sim I(0)$ y $Y_t \sim I(1)$, entonces $Z_t = (X_t + Y_t) = I(0)$
2. Una combinación lineal de una serie I (d) es también I (d).
3. Cuando $X_t \sim I(d_1)$ y $Y_t \sim I(d_2)$, decimos que $Z_t = (aX_t + bY_t) \sim I(d_2)$, donde: $d_1 \leq d_2$ y a y b son constantes.
4. Si $X_t \sim I(d)$ y $Y_t \sim I(d)$, entonces $Z_t = (aX_t + bY_t) \sim I(d^*)$ señalando que $d^* \sim d$, pero que en algunos casos $d^* \leq d$, como sucede en la cointegración.

XI.3.3.-METODOLOGIA DE BOX & JENKINS APLICADA EN LA PREDICCCION DE SERIES TEMPORALES.

XI.3.3.1.- Conceptos básicos necesarios para entender esta metodología

XI.*3.3.1.1.-Proceso estocástico*:

Es el registro en el tiempo de los valores de una variable o serie estadística de interés para el investigador, los cuales se originan aleatoriamente. En economía puede ser la producción de ladrillos, de tornillos, agujetas, etc., que genera una maquina en un lapso determinado. En medicina pueden ser los latidos del corazón, el pulso de un paciente; en el clima, los registros de la temperatura diaria o de la lluvia en un tiempo previamente establecido. *Todo este proceso de registro de las variaciones o fluctuaciones de la variable o serie temporal se hace con el fin de estudiarlas e interpretarlas en el contexto de la estabilidad o estacionariedad que es necesario que las caracterice para hacer predicciones apropiadas en el futuro*. En economía, finanzas y en los negocios esto último es muy importante para predecir por ejemplo el PIB, las tasas de interés, el valor de las acciones de una empresa, etc.

Así, decimos que un proceso estocástico estacionarioo estable es el que tiene μ, σ^2 y covarianza constantes, i. e., *son invariantes respecto al tiempo*. En opinión de Pulido et al 2000,168): “si presenta un componente tendencial (no estacionariedad en media) o si la dispersión no es homogénea en el tiempo (no estacionariedad en varianza)”, el proceso no es estable, se dice que no es estacionario.

También se le llama: a) Proceso estocástico débilmente estacionario; b) Estacionario covariante; c) Estacionario de segundo orden; d) Proceso estocástico en sentido amplio. Ejemplo, si Y_t se desplaza a Y_{t+m}, entonces la μ, σ^2 y covarianza de Y_t es la misma que para Y_{t+m}, donde “m” es un lapso de tiempo.

XI.*3.3.1.2*.-Importancia de la estacionariedad en la identificación de tendencias.

Derivado de lo anterior, podemos decir que la estacionariedad permite distinguir si un proceso estocástico es o no estacionario, es decir, estable; lo anterior permite saber si la tendencia (evolución lenta de la variable en el largo plazo) ES DETERMINISTA O ESTOCÁSTICA. Al respecto decimos que es determinista cuando sus valores se pueden obtener con precisión. Si no es así decimos que dicha tendencia es estocástica, i. e., es predecible con el uso de la probabilidad.

Luego la distinción entre procesos estocásticos estacionarios y no estacionarios es muy importante porque nos permite saber si la tendencia es:

a).- Tendencia Estacionaria (TE); o

b).- Diferencia Estacionaria (DE).

Esta identificación nos permitirá seleccionar el método para transformar una serie no estacionaria en estacionaria.

XI.*3.3.1.2.1.-Ruido blanco*

Proceso puramente aleatorio o de ruido blanco es un proceso estocástico (serie de tiempo) ESPECIAL, es aquel cuya μ = 0; σ^2= 1son constantes y él no está serialmente correlacionada. Un ejemplo es U_t: Término de error, que es un proceso con ruido blanco, i. e., sus términos no están autocorrelacionados entre si, por eso decimos que en este proceso los términos se distribuyen de manera independiente y es idéntico en su comportamiento a la distribución normal porque tiene $\mu = 0$ y $\sigma^2 = 1$.

XI.3.3.1.2.2.-Modelo de Caminata Aleatoria, MCA

Es una particularidad de un proceso estocástico *no estacionario*. Decimos que hay dos tipos de MCA, que son:

a).- MCA sin variaciones, cuando la ecuación no tiene ordenada al origen y b).- Con variaciones, cuando si la tiene.

Para el primer caso MCA se puede expresar diciendo que si Y_t es una serie temporal, decimos que $Y_t = Y_{t-1} + U_t$ es una MCA pura,

Donde U_t sí es estacionaria. Se llama pura porque el valor de Y_t es igual a su valor de "ayer" (Y_{t-1}) más un choque aleatorio (U_t); éste último puede interpretarse como un suceso inesperado (innovación, algo imprevisto), que está fuera de control o manipulación alguna y que *se comporta como una distribución normal estándar*, que tiene $\mu = 0$ y $\sigma^2 = 1$, motivo por el cual se le conoce como un proceso con ruido blanco, por lo que es una serie estacionaria.

Así, dicho en otras palabras, si $Y_t = Y_{t-1} + U_t$, en cuyo caso U_t es una serie aleatoria con media, variancia y covariancia invariantes en el tiempo, además de que U_t *no está serialmente correlacionada.*

El Modelo de caminata aleatoria: MCA, es una serie no estacionaria y es un ejemplo DEL PROCESO DE RAÍZ UNITARIA.

XI.3.3.1.2.3.- Raíz unitaria

Si escribimos MCA como $Y_t = \rho Y_{t-1} + U_t$; donde $-1 \leq \rho \leq +1$; y ρ es el coeficiente de autocorrelación puro o serial en los términos de error U_t. Así si $\rho = 1$, entonces $Y_t = Y_{t-1}$, es una MAC sin variaciones y se tiene lo que se conoce como PROBLEMA DE RAÍZ UNITARIA IGUAL A SERIE NO ESTACIONARIA y escribimos $Y_t – Y_{t-1} = U_t$.

Ahora, si L = operador de rezago tal que $LY_t = Y_{t-1}$ y también $L^2Y_t = Y_{t-2}$, por lo que la ecuación inicial de MCA que es $Y_t = \rho Y_{t-1} + U_t$, puede escribirse como:

$Y_t = \rho Y_{t-1} + U_t$

Ahora bien, si ρ: coeficiente de autocorrelación de primer orden, cuyo valor oscila entre 0 y ± 1 y si $\rho = 1$, entonces:

$Y_t – Y_{t-1} = U_t$ $\quad$ U_t es ruido blanco

Factorizando:

$Y_t (1 – L) = U_t$.

Si el término raíz unitaria se refiere a la raíz del polinomio en el operador de rezago, entonces sí (1 – L) = 0, luego L =1, de ahí el nombre de RAÍZ UNITARIA.

Luego si MCA tiene raíz unitaria, ello significa que la variable no es estacionaria , que es caminata aleatoria, que no es rudo blanco y que existe autocorrelación serial e indicio de que la correlación sea por casualidad y no por causalidad (Suriñac, 1999). Se elimina al hacer estacionaria La serie temporal DIFERENCIANDOLA, ergo, $\Delta Y_t = (Y_t - Y_{t-1}) = U_t$ que como se recordará es ruido blanco o estacionaria, en otras palabras, no está correlacionada serialmente y se distribuye normalmente.

XI.*3.3.1. 2.4.- Diferenciación*

La transformación de MCA mediante el método de diferenciación tiene el siguiente fundamento:

Si tenemos el modelo $Y_t = \beta_1 + \beta_2 X_t + U_t$, donde Y = Gasto en consumo; X = Ingreso. Como la ecuación anterior es válida para cada periodo, también lo es para el periodo anterior (t-1) y se escribe así:

$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + U_{t-1}$.

Decimos que a Y_{t-1}, X_{t-1} y U_{t-1} se les llama VALORES REZAGADOS DE Y, X, U, respectivamente. En este caso específico las tres variables están rezagadas un periodo y a la ecuación $Y_t = \beta_1 + \beta_2 X_t + U_t$ se le llama ecuación FORMA DE NIVEL.

Si restamos la segunda de la primera ecuación tenemos:

$\Delta Y_t = \beta_2 \Delta X_t + \Delta U_t$; llamada ecuación en primera diferencia.

A "Δ" se le conoce como primer operador de diferencia, el cual expresa que se obtuvieron diferencias sucesivas de Y, X y U, es decir:

$\Delta Y_t = (Y_t - Y_{t-1})$

$\Delta X_t = (X_t - X_{t-1})$

$\Delta U_t = (U_t - U_{t-1})$

Hipotéticamente, si n = 10, numéricamente lo anterior se ve para Y, diferenciada tres veces, (Rojas, 2009) así:

Tabla 11.3.3.1. 2.4.1 Datos

Observación	**Y**	**Primera diferencia (DY)**	**Segunda diferencia (D2Y)**	**Tercera diferencia (D3Y)**
1	5	-	-	-
2	6	1	-	-
3	7	1	0	-

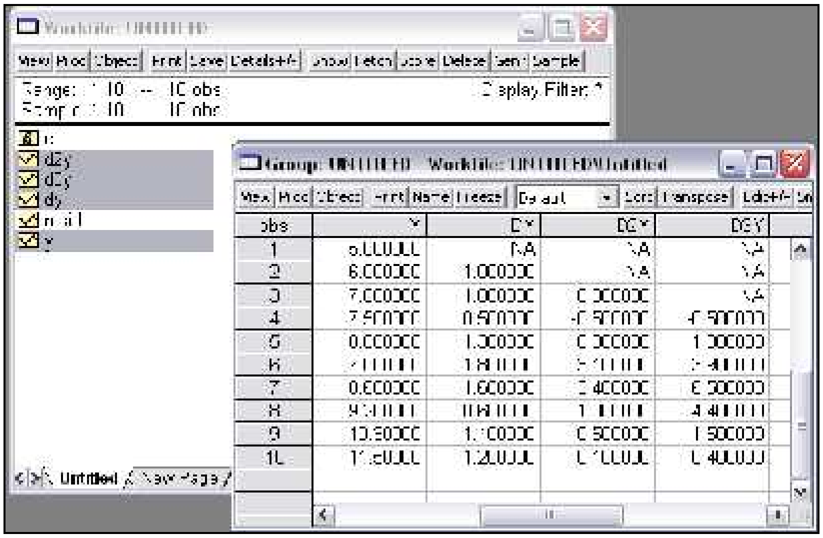
4	7.5	0.5	-0.5	-0.5
5	8.8	1.3	0.8	1.3
6	7.0	-1.8	-3.1	-3.9
7	8.6	1.6	3.4	6.5
8	9.2	0.6	-1.0	-4.4
9	10.3	1.1	0.5	1.5
10	11.5	1.2	0.1	-0.4

Así, dados los valores de la variable Y, **la primera diferencia** de la serie Y consiste en restar el valor de la observación 2 al valor de la observación 1, 6 – 5 = 1, luego, al valor de la observación 3 se le resta el valor de la observación 2, 7 – 6 = 1, luego al valor de la observación 4 se le resta el valor de la observación 3, 7.5 – 7 = 0.5, etc., con lo cual para el caso de la primera diferencia se pierde una sola observación obteniéndose la serie "DY" o la primera diferencia de Y. Ahora, para **la segunda diferencia** (D2Y) se parte de la serie de la primera diferencia (DY) y no de la serie original Y, así, a la observación 3 de la serie DY se le resta el valor de la observación 2 de dicha serie, es decir, 1 – 1 = 0, al valor de la observación 4 de DY se le resta el valor de la observación 3 de DY, 0.5 – 1 = -0.5, al valor de la observación 5 de la serie DY se le resta el valor de la observación 4 de dicha serie, 1.3 – 0.5 = 0.8, etc., con lo cual para el caso de la segunda diferencia (D2Y) se habrán perdido dos observaciones, a saber, una de la primera diferenciación y otra de la segunda. Para la **tercera diferencia** (D3Y) se parte de la serie obtenida a partir de la segunda diferencia (D2Y), así, al valor de la observación 4 de la serie D2Y se le resta el valor de la observación 3 de dicha serie, -0.5 – 0 = -0.5, al valor de la observación 5 de D2Y se le resta el valor de la observación 4 de la misma aserie, 0.8 – (-0.5) = 0.8 + 0.5 = 1.3, al valor de la observación 6 de D2Y se le resta el valor de la observación 5 de la serie en segunda diferencia, -3.1 – 0.8 = -3.9, etc., con lo cual se habrá obtenido la serie en tercera diferencia (D3Y), y así sucesivamente si se requiere de un orden de diferenciación mayor.

Esta diferenciación utilizando EViews, se realiza con el siguiente procedimiento:

1. Una vez generado el archivo de trabajo y con las series capturadas, generamos las series en diferencias, así vamos a: "Quick/Generete Series y escribimos: dy=d(y), donde "d" indica que se trata de la primera diferencia de la serie Y.
2. Para la segunda diferencia siguiendo el mismo procedimiento, ahora escribimos: d2y=d(y,2), en donde el 2 indica la segunda diferencia de la serie Y; y para la tercera diferencia escribimos d3y=d(y,3), en donde 3 indica la tercera diferencia, etc.
3. Para visualizar la serie diferenciada (que no rezagada, véase definición en modelo dinámicos, párrafos atrás) seleccionamos las variables generadas en el archivo de trabajo (work file) y las abrimos como grupo (as group):

Grafico 11.3.3.1. 2.4.1 Obtencion



Por otra parte, debe aclararse que una serie no estacionaria se puede estudiar y usarse para pronosticar solo para el periodo de tiempo bajo consideración, en tanto que con una estacionaria se puede generalizar para todos los periodos, puesto que sus variaciones se ajustan en torno a su media y su varianza, de ahí que se recomiende siempre obtenerla**.**

Al respecto, dicho en otras palabras, es necesario “descontaminar” la serie de la variable endógena de los elementos irregulares de las estacionalidades que inducen tendencias equivocadas. Esta descontaminación se logra con diversos métodos dentro de los que destacan el de las MA, AR, etc., que veremos más adelante. Esta “limpia o descontaminación” debe hacerse porque hay datos que *transportan* componentes indeseables que generan “ruido” en la predicción. Luego hay que limpiar o descontaminar la serie depurándola, eliminándole los componentes irregulares estacionales y quitar la tendencia a que nos están conduciendo equivocadamente. Esta acción de descontaminación también se conoce como *filtrado*, su resultado es que al hacerlo producen ahora una serie con ruido blanco: sin efectos negativos como por ejemplo la autocorrelación.

Los “filtros” de mayor uso son, como arriba se indicó: los autorregresivos o diferenciados, (AR: auto regression) y los aditivos o de medias móviles, (MA:moving averages)

Dentro de los primeros (Gabriel Tapia Gómez, et al, 2002) el filtro univariante más sencillo, AR, es aquel en que el operador diferencia, digamos que obtiene la primera diferencia ordinaria y se denota así: AR(1).Con este procedimiento se eliminan las frecuencias bajas pero se acentúan las más altas. Comenta que en series suaves permite obtener estimaciones de los componentes estacional e irregular y aproximar la tasa de crecimiento de la serie original cuando se filtra el logaritmo y, por consiguiente se deduce que se utiliza como indicador del ciclo cuando las series tienen comportamientos suaves. En este contexto es interesante decir que existen otros filtros autoregresivos como el AR de orden mayor, que consiste en obtener la diferencia de orden estacional, con lo que se eliminan las frecuencias bajas y los ciclos armónicos de la frecuencia fraccional correspondiente.

El segundo filtro, MA, se utiliza para extraer componentes en análisis del ciclo, ya que "dejan pasar intacta la información contenida en determinada banda de frecuencias mientras que eliminan o acentúan las restantes". Con este método se desestacionaliza extrayendo el componente ciclo tendencia.

Derivado de lo anterior podemos decir que con los filtros MA y AR, sucesivamente, se tendrán como resultados correspondientes, la atenuación de altas frecuencias y la atenuación de bajas frecuencias, resultando un pico en la función de transferencia de la serie filtrada y una atenuación de determinadas frecuencias intermedias. En opinión del investigador Gabriel Tapia, et al, los filtros más usados son los de Hodrick & Prescott (1980), Prescott (1986) y los de Henderson.

XI.*3.3.1.2.5.*-Comentarios: Esta metodología ideada por Box & Jenkins para estacionarizar una serie ha sido cuestionada porque sus críticos comentan que al estar "diferenciando" se pierde información de la serie, dado que el problema se resuelve sólo para los datos que se usan para obtener las diferencias con las que se gesta una nueva serie supuestamente con datos que fluctúan menos en el tiempo. Si bien es cierto lo anterior, dicha solución resuelve el problema de que las fluctuaciones afectan la estabilidad de la variable en el tiempo. Como se verá más adelante para superar esta limitante (aprovechar toda la información) fue ideada la metodología "Corrección del error" que, al conjugarse con la metodología de cointegración (para el largo plazo) se logra obtener la estabilidad de la variable, lo cual facilita hacer planeación en cualquier horizonte temporal de interés para el investigador.

Finalmente, se acostumbra escribir $\Delta Y_t = \beta_2 \Delta X_t + V_t$; donde V_t es ΔU_t. Claramente se intuye que ΔY y ΔX representan *cambios* (diferencias) en tanto que Y_{t-1} y X_{t-1} representan variables en *forma de nivel*.

Luego una ecuación en forma de nivel expresa las relacionesentre los valores originales de las variables (Y, X), en tanto que una ecuación en primera diferencia expresa cambios que por ejemplo pueden ser tasas de crecimiento, de manera que esta última ecuación expresa relaciones de Y y X en forma de crecimiento.

Por otra parte, conviene observar que si U_t: Término de error calculado con MCO y sí se prueba que no está autocorrelacionada (que es estacionaria), ***se debe tener cuidado*** porque también se puede probar que V_t: Término de error de su primera diferencia, sí está autocorrelacionada (no es estacionaria).

Los modelos con $\Delta Y_t = \beta_2 \Delta X_t + V_t$ se conocen como modelos dinámicos de regresión, mismos que se caracterizan por incluir variables independientes rezagadas (ΔX_t).

XI.*3.3.1.2.6.*-Conclusiones

La primer conclusión que podemos sacar de transformar el modelo original ($Y_t = \beta_1 + \beta_2 X_t + U_t$) en $\Delta Y_t = \beta_2 \Delta X_t + V_t$ es que a veces puede inducirse la autocorrelación como consecuencia de haber transformado el modelo original. Otra conclusión es que cuando existe autocorrelación perfecta, se dice que $\rho = 1$ y L = 1, que precisamente por eso existe lo que se llama *raíz unitaria*, la cual indica que la serie es no estacionaria. Otra conclusión es que en la práctica se acostumbra decir en general que son sinónimos los términos estacionariedad, ruido blanco y no autocorrelación; igualmente, que son sinónimo los conceptos de no estacionariedad, caminata aleatoria y autocorrelación. Finalmente, debe agregarse que cuando $|\rho| \leq 1$ está demostrado que la serie temporal Y_t es estacionaria.

XI.*3.3.1.2.7.-* La regresión espuria.

Es una relación que se da *por casualidad y no por causalidad* (Suriñac, et al, 1999) entre 2 variables en estudio, cuando se estima con MCO, simplemente porque ambas varían en el tiempo y son asociadas por el investigador. Suele suceder que cuando se hace la regresión $Y_t = f(X_t)$ y si son variables o procesos estocásticos no correlacionados, el investigador espera que su R^2 sea bajo; sin embargo, <u>suele</u> aparecer con un R^2 alto, ello se denomina regresión espuria o falsa que Yule (1926) llamo "regresión extraña" entre Y_t y X_t, Al respecto, Granger y Newbold (1974) indican que *cuando R^2 es alta y d de Durbin-Watson es baja*, ello es indicio de que **existe una relación espuria entre Y_t y X_t**. Esta extraña relación puede eliminarse aplicando el método de diferenciación; ejemplo, cuando Y_t e X_t son series <u>no</u> estacionarias, su R^2 es alto y d es baja, se dice que hay regresión espuria; para que desaparezca se hace la regresión de las primeras diferencias de Y_t (= ΔY_t) sobre las primeras diferencias de X_t (= ΔX_t), mismas que ahora son estacionarias y determinan que su R^2 se acerque a cero y que d de Durbin-Watson a 2, i. e., indican que ha desaparecido la relación falsa entre Y_t e X_t. Ella aparece con más frecuencia cuando Y_t e X_t exhiben tendencias estocásticas y que por consiguiente son I (1).

Con estas referencias conceptuales a continuación expondremos el

XI.3.3.1.2.8.- Procedimiento para obtener el modelo ARIMA

Pérez (2007) sugiere los siguientes pasos para obtener el modelo ARIMA, cuya modelización se deriva después de haber hechos pruebas y arribar a los valores adecuado que se obtengan para (p,d,q); en que **p** se refiere al número de autorregresiones (AR) a realizar, **d** al número de diferenciaciones en la integrabilidad (I) y **q**, al número de aplicaciones de medias móviles (MA, en inglés). Dichos pasos son:

1. trabajar con n≥50 datos;

2. Graficar Y_t; a veces se usan medias y varianzas por subperíodos para identificar por tramos mejor la posible “estacionariedad” de la serie temporal Y_t.

3. Transformar Y_t logarítmicamente para alisar su variabilidad procurando su estacionariedad en media y en varianza; lo recomendable es ensayar con sus datos originales y en logaritmos

4. Si se identifica alguna tendencia, la lineal se corrige tomando primeras diferencias (d), que se indica así: I(1), que para bien o para mal es el caso más frecuente; cuando la tendencia es no lineal, su corrección requiere hacer quizás dos o más diferencias;

5. Identificación y formulación del modelo: se hace corrigiendo los procesos Ar (p) y MA(q). Técnicamente lo primero se realiza analizando las funciones de autocorrelación (FAC) y autocorrelación parcial (FACP); por lo general la formulación se prueba con AR(1), AR(2) y MA(1), MA(2) y se culmina con ARMA(1,1) que toma en cuenta la parte regular (AR) y la estacional (MA);

6. Una vez identificado y formulado el modelo, el siguiente paso es

La estimación de los parámetros;

7. Enseguida se verifica la bondad de ajuste contrastando su validez conjunta con las estadísticas t, F, covarianza, R^2, $\sum U_t$, etcétera;

8. Se sugiere hacer el análisis detallado de los errores U_t. Se calculan y estudian las diferencias entre $Y_t - Y_{t-1} = U_t$; se deberá comprobar un comportamiento no sistemático de los errores U_t;

9. Una vez identificados los resultados satisfactorios en las fases anteriores, se procede a seleccionar el modelo que mejor se adapte a los datos de Y_t;

10. Predicción: ahora ya se está en condiciones de iniciar la predicción para el horizonte de planeación temporal deseado.

XI.3.3.1.2.8.1.1.-Características de los modelos ARIMA con valores apropiados de p,d, q:

1. Un modelo ARIMA, por ejemplo (0,d,0) es una serie temporal que se hace o convierte en estacionaria o ruido blanco (proceso puramente aleatorio) después de ser diferenciado d veces, diferenciación realizada con objeto de extraer las posibles fuentes de no estacionariedad (de variación), mismas que de no extraerse se expresan o manifiestan con media y varianza variantes en el periodo de tiempo de estudio de Y_t;
2. La transformación Box-Jenkins *consigue estabilizar la variación* de una serie temporal (serie estacionaria en media y en varianza) y aproxima su distribución a la normal: Yt tiende N(0,1), es decir con media cero y varianza 1. Al respecto,

conviene reiterar que la inducción de la estacionariedad en Y_t por lo general se hace mediante la diferenciación estacional, que puede ser de primer orden, segundo orden, etc. i.e, hasta que se consiga que sea ruido blanco o estacionaria.

XI.3.3.1.2.8.2.-Descripción detallada de la metodología

XI.3.3.1.2.8.2.1.-Referencias

Puesto que su aplicación se hace sobre series temporales, es conveniente reiterar que una serie de tiempo se define como el registro de una sucesión de valores de una variable económica, Y_t, en un periodo de tiempo de interés para el investigador. Como se indicó antes, conviene reiterar que la serie puede estar compuesta por: a) Tendencias; b) Variaciones estacionales; c) variaciones cíclicas y d) variaciones residuales (Pérez, 2007:537).

Al respecto, cabe mencionar que dentro de la estrategia didáctica para transmitir el conocimiento descrita al inicio de este libro, **que consiste en la reiteraciónsin abrumar al lector** de los conceptos cuando se juzgue conveniente, a continuación diremos que:

Tendencia: se entiende como un movimiento general a largo plazo de Y_t;

Variaciones estacionales: son oscilaciones que se producen o repiten en un lapso igual o inferior a un año calendario;

Variaciones cíclicas: Son oscilaciones observadas, reconocibles en un periodo mayor a un año, derivadas de repeticiones (ciclos) alternados;

Variaciones residuales o irregulares: así se denomina a los movimientos de Y_t con un comportamiento errático, generalmente no reconocible periódicamente, que son provocados por acciones singulares que afectan a Y_t esporádicamente.

Método de cálculo para identificar la tendencia:

I. Tendencia de $Y_{t;}$ para hallarla, identificarla y ajustarla matemáticamente a los valores reales se usan los métodos: a) de ajuste analítico; b) medias móviles y c) el de las diferencias.

Método de ajuste analítico, se pueden usar formas funcionales lineales: $Y_t = a+bt$; cuadráticas: $Y_t = a+bt+ct^2$; logarítmicas, exponenciales, etc.

La bondad de ajuste que más destaca por ejemplo en la forma funcional lineal, es que es útil porque muestra aumentos o disminuciones *constantes*(valor de b) en los valores temporales de la serie Y_t. En lo que atañe a la forma funcional logarítmica, es útil porque permite suavizar los datos de la serie cuando aumentan o disminuyen *rápido a abruptamente*; esta forma funcional puede usar valores positivos o negativos. La Semilogarítmica es una variante de la forma logarítmica.

La tendencia potencial: Se expresa con una línea curva cuya ecuación es $Z_t = at^b$ permite comparar datos que aumentan a un ritmo determinado, ejemplo: la velocidad con que se encauza un automóvil. Como se detecta, en este caso no es posible aplicar esta forma funcional si los datos incluyen el cero o son negativos.

La tendencia Exponencial: Su curva se describe con la ecuación Z_t =Exp(a+bt), misma que se aplica cuando los valores cambian (aumentan o disminuyen) a intervalos cada vez mayores. Como el caso anterior: no se usa cuando los datos toman el valor cero o negativos.

XI.3.3.1.2.9.- Metodología específica para la predicción

En línea con lo antes expuesto ahora debemos decir que la metodología de Box & Jenkins, que sirve de referencia básica para la Predicción de valores de las variables en estudio se realiza, entre otros, utilizando dos métodos muy conocidos, ellos son:

1. El método autorregresivo integrado de medias móviles (ARIMA); y

2. El método de autorregresión vectorial (VAR).

En este sentido conviene decir que, profundizando en el tema, existen cinco enfoques para la predicción económica que se hace de las series temporales, los cuales son: i) El método de alisamiento exponencial; ii) los modelos de regresión uniecuacionales; iii) Los modelos de ecuaciones simultáneas; iv) Los ARIMA y v) los VAR.

El método de alisamiento exponencial, con él es posible ajustar una curva ad-hoc a los datos históricos de cierta variable económica en estudio; destacan los métodos de Holt y Holt- Winthers. En la actualidad se usan poco porque los otros cuatro prácticamente los han sustituido.

Los modelos de regresión uniecuacionales se caracterizan porque a partir de los datos de las series temporales se puede estimar un modelo apropiado / Líneal, log-líneal (o no lineal) para hacer predicciones de la variable en el futuro.

Los modelos de ecuaciones simultáneas se construyen para predecir con información pasada de las variables incluidas en el modelo: endógenas y exógenas. Cayeron en desuso con las crisis petrolera y monetaria de la década de los 70´s del siglo XX; también, por la crítica de Robert E. Lucas, quien afirmaba que “Los parámetros estimados de un modelo econométrico dependen de la política prevaleciente en el momento en el que se estima y que cambiarán si hay un cambio de política”, como efectivamente sucedió con el cambio de política petrolera y monetaria de esa década.

Los Modelos ARIMA de predicción son uniecuacionales que hacen énfasis en "el análisis de las propiedades estocásticasde las series temporales". Se caracterizan porque la variable Y_t es explicada por valores *pasados o rezagados* de si misma, por medias móviles y por los términos estocásticos de error, U_t. Esta situación los hace diferentes de los modelos en que Y_t es explicada por n variables individualmente (X_1, X_2 , …….,X_n).

Los modelos ARIMA de mayor uso son **univariantes,** así llamados porque se construyen con una sola serie temporal, independientemente de que su análisis sea multivariable en ciertos casos. Fueron de gran uso en la década de los 70´s del siglo XX, prácticamente sustituyeron a los modelos multiecuacionales, **MES**, que se basan en un Sistema de Ecuaciones Simultámeas, **SES**.

*Los modelos VAR*se parecen a los **MES** en el sentido de que incluyen varias variables endógenas en su modelización; sin embargo, <u>se diferencian en que cada variable endógena es explicada por sus valores pasados o rezagados, así como por los valores pasados o rezagados del *resto* de las variables endógenas consideradas en el modelo</u>. Como puede observarse <u>no utilizan variables exógenas</u> para construir el modelo de predicción.

XI.3.3.1.2.9.1.- Modelos ARIMA

Analizando los dos últimos modelos aquí expuestos, empezaremos por preguntarnos ¿**Porqué se llaman modelos ARIMA?**

Así se le conoce porque están integrados por tres componentes: La autorregresión (AR), la Integrabilidad (I), y las Medias Móviles (MA).

a).-. Procesos Autorregresivos (AR)

Como ya se vio cuando se verificó si habían sido violados los supuestos básicos de MCO, concretamente la independencia o ausencia de correlación entre los términos de una serie, ahora conectando esa temática con el concepto de estacionariedad de la serie temporal denotada por Y_t, diremos que cuando ésta es estacionaria se puede modelar de maneras diferentes, una de ellas es la AR.

Así, si Y_t ahora la modelamos como: $(Y_t - \omega) = a_1 (Y_{t-1} - \omega) + U_t$ en que ω representa la media aritmética de Y y U_t representa el término de error aleatorio no correlacionado, motivo por el cual se dice que éste último es ruido blanco, dado que es estacionario con momento uno (media constante) y momento dos (varianza constante). En este caso concreto consideramos que Y_t sigue un proceso estocástico autorregresivo de primer orden, que se denota **AR(1)**:

Al respecto, por definición decimos que cada uno de los valores de Y en el tiempo t <u>depende de su valor en el periodo anterior y de un término aleatorio</u>. *En esta ecuación los valores de Y son sus desviaciones o variaciones con respecto a ω*. Esta modelización tiene una connotación muy especial e importante dado que indica que "el valor de pronóstico de Y_t en el periodo t es simplemente alguna proporción (=a_1) de su valor en el periodo (t-1) *"más un choque"* o perturbación en el tiempo (Gujarati, 2004: 812).

Por otra parte, si ahora el modelo se construye con la siguiente ecuación:

$$(Y_t - \omega) = a_1(Y_{t-1} - \omega) + a_2(Y_{t-2} - \omega) + U_t$$

En él ahora se indica que Y_t se comporta como un proceso autorregresivo de segundo orden, escrito asi: **AR(2).** Ello significa que ahora el valor de Y_t en el lapso t depende de sus valores en los *dos* periodos anteriores y de un término aleatorio. En general si modelamos los valores de Y en torno a sus desviaciones con respecto a su media aritmética ω, dicho modelo se expresa con la siguiente ecuación así:

$$(Y_t - \omega) = a_1(Y_{t-1} - \omega) + a_2(Y_{t-2} - \omega) + \ldots\ldots\ldots + a_p(Y_{t-p} - \omega) + U_t$$

A este proceso lo llamaremos proceso autorregresivo de orden **p**, y por motivos prácticos lo escribimos como **AR(p).**

El lector puede observar que en estas tres modelizaciones únicamente se consideran los valores actual (Y_t) y anteriores (Y_{t-p}) de Y, motivo por el cual Box & Jenkins consideran que "los datos hablan por sí mismos".

b).-Proceso estocástico Integrado (I)

Se llama integración al proceso mediante el cual si Y_t no es estacionaria se "diferencia" para transformarla en estacionaria. Ejemplo, si Y_t se diferencia una vez, *se le llama proceso integrado de orden 1, el cual por motivos prácticos se escribe* ***I(1)***. Si Y_t tiene que diferenciarse dos veces para transformarla en estacionaria, se le conoce como proceso integrado de orden dos, **I(2)**. Así, como lo hicimos con AR, si generalizamos diciendo que si Y_t es no estacionaria y ésta debe diferenciarse **d** veces para transformarla en estacionaria, se dice que la serie temporal Y_t es un proceso estocástico integrado de orden d, **I(d)**.

Como podrá intuirse, si una serie temporal es inicialmente estacionaria, de manera que no es necesario diferenciarla, se le conoce como proceso estocástico integrado de orden cero y que se denota como $\mathbf{Y_t \sim I(0)}$.

De lo anterior deducimos que son sinónimos los conceptos de "serie de tiempo integrada" de orden cero", "serie de tiempo estacionaria" y "proceso estocástico integrado de orden cero" .

Como señala Gujarati (2004: 779) la mayor parte de las series temporales económicas son I(1), es decir, son no estacionarias, por lo que es necesario diferéncialas por lo menos una vez para convertirlas en estacionarias.

c) **Procesos de Medias Móviles (MA)**

Se generan calculando las medias móviles de los datos originales de las series de tiempo. Cada uno de los puntos de un proceso MA es el promedio del número de términos que el investigador estime conveniente usar para calcularla; puede ser de los dos primeros términos, puede ser con el segundo y el tercero, etc.

Por ejemplo, si se tiene un conjunto de 100 datos el primer valor de la serie de medias móviles podría ser el promedio de los primeros 25 términos, luego el promedio de los términos 2 al 26, el tercer elemento de los términos 3 al 27 y así, hasta por último el promedio de los últimos 25 números del 76 al 100.

Una serie de medias móviles puede ser calculada para cualquier serie temporal. Se usa por ejemplo para calcular una demanda estable, sin tendencia y sin estacionalidad; suaviza las fluctuaciones de plazos cortos, resaltando así las tendencias o ciclos de plazos largos."

Dentro de la metodología de Box & Jenkins, así se les llama a estos procesos porque son una combinación lineal de los términos de error, U_t, con ruido blanco y una media móvil. Para ilustrarlo suponga usted que modelamos Y_t con la siguiente ecuación:

$Y_t = \mu + \beta_0 U_t + \beta_1 U_{t-1}$ en la que μ es considerada una constante y U_t tiene el mismo significado de antes, i.e., es el término de error estocástico o aleatorio con ruido blanco y por ello es estacionario.

Esta ecuación tiene la siguiente *interpretación*: el valor de Y_t en el lapso t es igual a una constante (μ) más un promedio móvil de U de sus términos presente (U_t) y anterior (U_{t-1}). Luego en este caso se dice que Y_t sigue un proceso de media móvil de primer orden, que por motivos prácticos se escribe como **MA(1)**.

En cambio sí señalamos que Y_t se expresa con el modelo cuya ecuación es:

$$Y_t = \mu + \beta_0 U_t + \beta_1 U_{t-1} + \beta_2 U_{t-2}$$

En cuyo caso Y_t sigue un proceso de media móvil de segundo orden,**MA(2).** Así, generalizando si:

$$Y_t = \mu + \beta_0 U_t + \beta_1 U_{t-1} + \beta_2 U_{t-2} + \ldots\ldots\ldots\ldots + \beta_q U_{t-q}$$

Modelo al que se le denomina proceso de media móvil de orden **q**, que escribimos como **MA(q).**

XI.3.3.1.2.9.1.-Evaluación de los modelos ARIMA

¿Cómo saber si el modelo ARIMA es apropiado estadísticamente para hacer predicciones?

Para ello debemos seguir los siguientes pasos:

1.- Del modelo ARIMA obtener los residuos U_t;

2.- Obtener los FAC y FACP, graficarlos y si vemos que ningún de los °© o °©©=se salen de los límites del intervalo de confianza que hemos construido en torno a ellos con cierto nivel de significación (α), y

3.- Si la probabilidad de Q de (BP) y (LB) de Ljuang-Box, es mayor que digamos α=5%, entonces

4.- Decimos que los *residuos estimados* de U_t son puramente aleatorios, que son independientes, que no hay autocorrelación, etc., luego

5.- No es necesario buscar otro modelo ARIMA para predecir.

XI.3.3.1.2.9.1.2- Resumen de modelos ARIMA

Para ilustrar este tipo de modelos antes es conveniente reiterar la frase de Box & Jenkins de "permitir que la información hable por si misma", la cual yo interpreto como la observación que debemos hacer sobre la naturaleza y comportamiento de los datos de la serie temporal Y_t, con objeto de adaptarles un modelo apropiado que describa su trayectoria futura (predicción) con la exactitud que sea posible.

Al respecto, parte de esa exactitud se logrará cuando Y_t sea una serie temporal estacionaria. En este contexto es conveniente decir que en esta exposición *se hace mención a las series temporales débilmente estacionarias,* mismas que se caracterizan por ser estacionarias en su media (momento uno) y en su varianza (momento dos), que son constantes y "su covarianza es invariante en el tiempo".

Ahora bien, también recordando que una serie temporal Y_t "es integrada" cuando "es no estacionaria", deducimos pues que ambos conceptos son sinónimos; y sabiendo que Y_t es integrada de orden uno, I(1), si sacamos su primera diferencia la hacemos estacionaria porque la hemos transformado en I(0). Con este razonamiento, si Y_t es I(2), su segunda diferencia la transforma en I(0), i.e. en estacionaria; de manera que en general decimos que si Y_t es I(d), al transformarla mediante "d" diferenciaciones se arriba a una serie estacionaria de orden cero, I(0).

En esta tesitura, si para hacer estacionaria una serie temporal Y_t es necesario diferenciarla d veces, y enseguida se le aplica a Y_t el modelo ARMA (p, q), ahora decimos que Y_t **es ARIMA (p, d, q)** y se le conoce como una serie temporal *autorregresiva integrada de media móvil,* en la que precisando decimos que:

p: número de términos autorregresivos;

d: número de veces que Y_t se diferencia para hacerla estacionaria; y

q: número de términos de media móvil.

Ejemplo numérico: si ARIMA (3,2,4) ello implica que la serie Y_t debe diferenciarse dos veces (d=2) para hacerla estacionaria; de manera que esta serie ya transformada en estacionaria, ahora puede modelarse como un proceso (3,4) en otras palabras, que tenga tres términos AR y cuatro MA.

Como se infiere si d=0, ARIMA será (p,0,q) que simplificando por motivos prácticos lo escribimos así: ARIMA (p,q). Ahora bien, conduciéndonos casuísticamente, si por ejemplo ARIMA (p,0,0) ello indica que este proceso estocástico es un proceso estacionario AR (p) *puro* porque d=0 y q=0; luego si un proceso ARIMA es (0,0, q) ello indica que es un proceso MA(q) *puro* porque p= 0 y d=0. De estos ejemplos casuísticos concluimos que los valores que tomen las literales p, d y q nos orientarán sobre el proceso ARIMA a modelar.

Lo anterior es muy importante porque en la medida que construyamos un modelo que genere apropiadamente <u>la información de pronóstico</u> que buscamos, estaremos haciendo buenas predicciones porque con esta metodología sustentamos que Y_t es estacionaria, tal que μ, σ^2 y covariancia son invariantes en el tiempo. Como ejemplo ilustrativo puede verse el número dos en “los ejercicios”.

XI.3.3.1.2.9.2.- Modelos ARMA: Combinación de los procesos estocásticos autorregresivos AR y MA.

Se dice que es una combinación porque el modelo consta de dos partes, la parte autorregresiva (AR) y la de media móvil (MA). Se le llama modelo ARMA (*p*,*q*), donde *p* es el orden de la parte autorregresiva y *q* es el orden de la parte de media móvil.

Esta modelización se hace cuando se considera que Y_t *para propósitos de pronóstico*, tenga características (comportamiento en el tiempo) tanto de AR como de MA; llamándosele en este caso modelo ARMA y se dice que muestra un proceso **ARMA (1,1)** cuando se escribe así:

$Y_t = \theta + a_1Y_{t-1} + \beta_0U_t + \beta_1U_{t-1}$

La cual es una ecuación en la que existe una parte autorregresiva ($\theta + a_1Y_{t-1}$) y otra de media móvil de U ($\beta_0U_t + \beta_1U_{t-1}$) y en donde θ es una constante.

Generalizando AR y MA, ahora podemos decir que en un modelo ARMA (p,q) hay p términos autorregresivos y q términos de media móvil.

XI.3.3.2.-Relación secuenciada de actividades a desarrollar para aplicar la Metodología de Box & Jenkins (BJ).

Derivado de la exposición previa en que se amplió el tema es que ahora los pasos sugeridos por Pérez (2007) debemos complementarlos diciendo y reiterando que para los modelos ARIMA inicialmente su metodología pretende indicar los valores que deben tomar p,d y q, para posteriormente, modelar apropiadamente el comportamiento de la serie temporal Y_t , de manera que produzca estimaciones robustas estadísticamente en el futuro visualizado por el investigador (corto, mediano y largo plazo).

Así pues, los pasos para modelar apropiadamente el comportamiento futuro de la serie temporal $Y_{t,}$, partiendo del punto 5 antes descrito por Pérez (ídem), son los siguientes:

1. Identificación del modelo: determinación de los valores apropiados de p, d y q. Para ello se sugiere utilizar FAC, FACP y sus correlogramas correspondientes.

2. Obtención de los parámetros del modelo seleccionado, corriendo la ecuación de regresión, por ejemplo con MCO.

3. Probar estadísticamente el modelo: la significación estadística de U_t para constatar que es ruido blanco.

4. Si se verifica que U_t es ruido blanco, la ecuación del modelo está lista para hacer predicciones de Y_t; y

5. **Si en cambio se observa que no es ruido blanco (es no estacionaria), el investigador debe regresar al punto 1 y determinar nuevamente p,d y q. , luego, seguir nuevamente los pasos 2 y 3.**

6. Predicción: correr la ecuación de Y_t para los horizontes temporales de interés del investigador (corto, mediano y largo plazo).

A continuación se describe la metodología a utilizar en cada uno de los pasos antes descritos:

Identificación del modelo

Para la identificación del modelo con los valores apropiados de p,d y q, se deben de calcular FAC: función de Auto Correlación; FACP: función de Auto Correlación Parcial; luego, graficarlos según la longitud del rezago en sus correlogramas correspondientes.

Decimos que los términos de la serie, los FAC: , están autocorrelacionados cuando éstos a lo largo del tiempo influyen o se impactan unos a otros; en este caso sus valores están cercanos a uno. Mientras más larga sea el periodo de análisis de la serie, menor es el efecto de los anteriores sobre los actuales. En otras palabras, decimos que la autocorrelación cuantifica la magnitud de la correlación o relación entre cada valor de Y_t y sus valores desfasados 1, 2, 3, 4, …, n periodos.

Haciendo la analogía con el coeficiente de correlación que mide la relación entre dos variables (X, Y) ahora refiriéndonos a la autocorrelación entre Y_t y ella misma un periodo antes (Y_{t-1}) decimos que el coeficiente de autocorrelación (ρ) que existe entre ellas *es de primer orden* y se calcula así:

$$\rho_1 = \frac{\operatorname{cov}(Y_t, Y_{t-1})}{\sqrt{\operatorname{var}(Y_t) * \operatorname{var}(Y_{t-1})}}$$

En un proceso estocástico estacionario su varianza es constante en t, por lo que $\operatorname{var}(Y_t) = \operatorname{var}(Y_{t-1})$, de manera que podemos escribir ahora que:

$$\rho_1 = \frac{\operatorname{cov}(Y_t, Y_{t-1})}{\operatorname{var}(Y_t)}$$

Generalizando: En un desfazamiento de k periodos, el coeficiente de autocorrelación de orden k se obtiene así:

$$\rho_k = \frac{\operatorname{cov}(Y_t, Y_{t-k})}{\operatorname{var}(Y_t)}$$

Observe usted que sí k=0, entonces:

$$\rho_0 = \frac{\operatorname{var}(Y_t)}{\operatorname{var}(Y_t)} = 1$$

Interpretación: El valor de ρ_0=1 significa que hay autocorrelación perfecta y expresa en otras palabras que Y_t tiene raíz unitaria, es decir, es no estacionaria.

La autocorrelación parcial, FACP, se define de manera similar al *coeficiente de correlación parcial.* Así*,* decimos que si en un modelo de regresión múltiple con k variables, el k-iésimo coeficiente de correlación parcial expresa el efecto de la variable regresora X_k sobre la variable regresada Y_t , manteniendo constante el efecto del resto de las variables regresoras sobre Y_t .

Similarmente, con ese referente, la autocorrelación parcial expresa la correlación entre los valores de los datos de la serie temporal, que están separados k periodos de tiempo **“manteniendo constantes las correlaciones de los otros rezagos.** Algebraicamente: la autocorrelación parcial entre dos términos de la serie se indica como la correlación entre Y_t y Y_{t-1}, manteniendo constante el efecto de los otros ρ.

Así, cuando los valores de FAC (son altos, ello se interpreta como que existe una alta autocorrelación entre los términos de U_t en el periodo t; su expresión gráfica, que es el correlograma, muestra que la mayoría de las observaciones (aparece afuera de las bandas o límites de confianza establecidos digamos con α= 5%. Se concluye que la serie Y_t es no estacionaria, que es necesario aplicarle primeras diferencias para hacerla estacionaria para enseguida modelizarla con AR y MA. Así, *la observación de los (que caen dentro y fuera de los límites del intervalo de confianza ayuda a determinar la longitud de los rezagos necesarios para hacer estacionaria una serie temporal que inicialmente no lo es*.

Además de lo anterior, también se puede aplicar las pruebas Dickey-Fuller (DF) y Aumentada Dickey- Fuller (ADF) que se usan para identificar la existencia de la raíz unitaria en la serie temporal Y_t.

Obtención del modelo ARIMA.

Luego entonces, una vez que se ha conseguido la serie temporal estacionaria, enseguida procede la construcción de su modelo ARIMA, **¿Cómo lograrlo? La respuesta** es usando nuevamente FAC y FACP y sus correlogramas asociados a un número seleccionado de procesos ARMA, por ejemplo, AR(1), AR(2), AR(3) y MA(1), MA(2), MA(3), es decir ARIMA(1;1), ARIMA(2,2), ARIMA(3,3) hasta conseguir un buen modelo (ecuación) de Y_t . Lo anterior implica estar experimentando con diferentes valores de p y q. Como lo anterior es muy laborioso en la práctica, Gujarati (2004:818) sugiere *lineamientos* como los que se describen enseguida para hallar dicho modelo, basados en los *patrones teóricosgráficos* que siguen FAC y FACP, ellos son:

Tabla 11.3.3.2.1 Obtencion de modelo ARIMA

Tipo de modelo	**Patrón típico de FAC**	**Patrón Típico de FACP**
AR (p)	Disminuye drásticamente	Picos grandes
MA(q)	Picos grandes	Decrece
ARMA(p,q)	Decrece exponencialmente	Decrece exponencialmente

Comentarios: Los FAC y FACP de los procesos AR(p) y MA(q) observan patrones contrarios. Con esas referencias se hacen las regresiones hasta encontrar un ARIMA (p,d,q) en que coincidan los patrones de FAC y FACP, mismos que sirven para obtener el modelo ARIMA, donde una de sus características es que el valor de R^2 es bajo y que **d=2** de Durbin Watson, indicando que ya no hay autocorrelaciones entre los términos de la serie

XI.3.3.2.1-Pronóstico

Con los resultados obtenidos enseguida ya se corre el modelo ARIMA para el horizonte de planeación que tenga en mente el investigador. <u>Este concepto se desarrolla ilustrandolo con detalle en el **siguiente ejercicio** con Wviews 5, así como en el Capítulo 12, punto 12.4.</u>

XI.3.3.3.-Algunas consideraciones sobre la metodología de BJ

1. Debe tenerse cuidado con la ecuación a utilizar para predecir, es decir, que se construya con los valores reales (**niveles** de las variables), no con los cambios (**diferencias**);
2. Observar que los datos no sean estacionales; si existe, eliminarla;

3.-Si bien es cierto que la metodología *se aplicó a una serie temporal a la vez,* característica de esta metodología, ello no impide que *se extienda al estudio simultáneo de dos o más series temporales*, en cuyo caso se usa la siguiente metodología de modelos VAR:

<u>Con este referente teórico, a continuación se ilustra la aplicación de su metodología.</u>

XI.4.- EJERCICIOS CON EVIEWS

Ejercicio 1 APLICANDO LA METODOLOGIA DE BOX & JENKINS.

A. Aspectos básicos con aplicaciones a la economía mexicana (Rojas: 2009).

El desarrollo de una aplicación de tipo ARIMA a través de la metodología Box & Jenkins debe partir necesariamente de un proceso estocástico estacionario, es decir, $\hat{\rho}_k$ ~ N(0, 1/n). Por lo que el primer paso a realizar es el análisis de la estacionariedad de dicho proceso, prestando especial atención al comportamiento constante en media y varianza (Pulido y Pérez, 2000: 167). Pues, la metodología Box & Jenkins precisa que las series sean estacionarias, por lo que es importante comprobar esta situación antes de trabajar con ellas.

Pero, ¿Porqué las series de tiempo estacionarias son tan importantes? De acuerdo a Gujarati (2007: 772-773) porque, si una serie de tiempo es no estacionaria, se puede estudiar su comportamiento sólo durante el periodo bajo consideración. Por tanto, cada conjunto de datos perteneciente a la serie de tiempo corresponderá a un episodio particular. Como consecuencia, no puede generalizarse para otros periodos. Así, pues, para propósitos de pronóstico (predicción), tales series de tiempo (no estacionarias) tendrán un valor práctico insignificante.

I) Identificación de la estacionariedad de la serie.

Existen dos formas de verificar la estacionariedad del proceso o de las variables aleatorias, a saber:

1) El método informal (gráfico); y
2) El método formal (las pruebas de raíz unitaria).

Cabe destacar que para ilustrar la aplicación de la metodología ARIMA con Eviews vamos a trabajar con el precio de las acciones del Consorcio ARA, la cual cotiza en la Bolsa Mexicana de Valores, para el periodo que comprende desde el 2 de enero de 2008 hasta el 19 de febrero de 2009, es decir, una muestra de 285 datos.

1. Método informal.

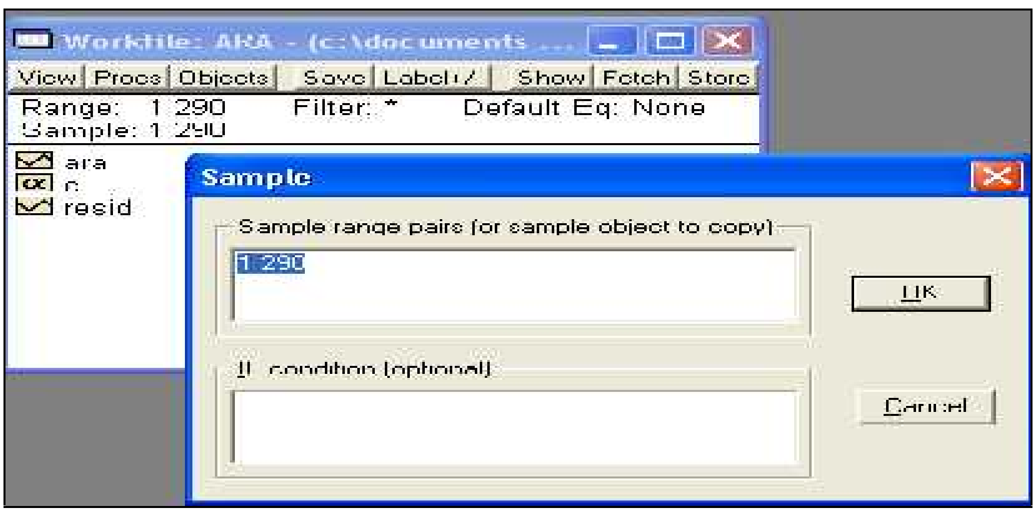
Como se trata del precio de acciones y dado que este mercado labora de lunes a viernes y se mantiene cerrado los sábados y domingos así como días festivos, lo lógico sería que seleccionáramos la frecuencia de datos diarios con semanas de 5 días (Daily [5 day weeks]). Desgraciadamente Eviews no reconoce los días festivos del calendario mexicano, por lo que nos marcará un tamaño de muestra más grande que la que poseemos. Sin embargo, en caso de que el lector desee realizar la prueba, en Eviews una vez que selecciono la frecuencia de los datos, en el rango debe escribir el periodo en el siguiente orden: mes:día:año, con el siguiente formato: mm:dd:aaaa. De tal suerte que, si el periodo de inicio es el 2 de enero de 2008, escribiremos en “Start date”: 01:02:2008 y si el periodo final de la muestra es el 19 de febrero de 2009 escribiremos en “End date”: 02:19:2009.

Para nuestros fines vamos a seleccionar frecuencia de datos irregulares (Undated or irregular) y dado que el tamaño de nuestro proceso estocástico es de 285 variables aleatorias en “Start date” vamos a escribir 1 y en “End date” 290 porque vamos a predecir el precio de las acciones de ARA para cinco días más. Así, 1 se referirá a la variable aleatoria del 2 de enero de 2008, 285 a la del 19 de febrero de 2009 y de la variable 286 a la 290 se referirá a la estimación de dicha variable, por lo que la variable 286 se refiere al precio del 20 de febrero, 287 a la del 23, 288 a la del 24, 289 a la del 25 y 290 a la del 26 de febrero de 2009.

Una vez generado el Workfile vamos al menú principal y ahí seleccionamos: Quick/Empty Group (Edit Series)/Ok. Lo cual nos desplegará una ventana de datos y en la cual capturaremos todas las variables aleatorias que componen el proceso estocástico. Mismo que denominaremos como "ARA". Dicho proceso aparecerá en la ventana del Workfile.

Una vez cargada la serie "ARA", para empezar a trabajar cambiamos el periodo muestral (Sample) en la ventana del Workfile dando doble click sobre el mismo, es decir, sobre "Sample". Este procedimiento desplegará una ventana denominada "Sample" que se observa de la siguiente forma:

Imagen 11.4.1 Comandos



En la ventana del Sample cambiamos el 290 por el 285. En la ventana del Workfile se observa ya el cambio realizado. Mientras que el Rango (Range) se mantiene inalterado. Lo cual presenta una estructura como esta:

Imagen 11.4.2 Comandos

a) Gráfica lineal y diagrama de dispersión.

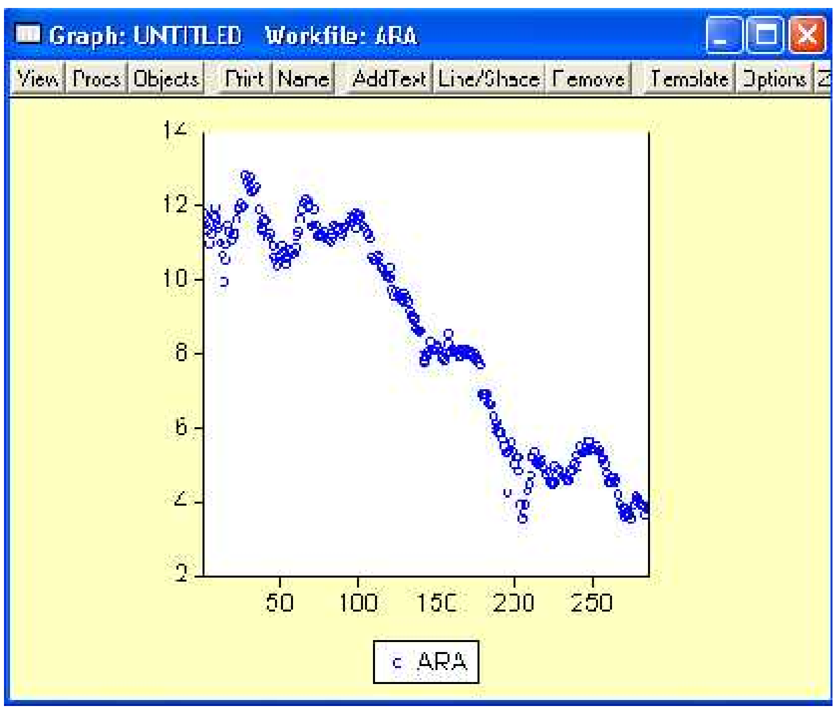
Hecho esto, abrimos la serie “ARA” dando doble click sobre ella en la ventana del workfile. Una vez abierta la ventana de la serie vamos a: View/Graph/Line/Ok. Lo que nos desplegará una gráfica como la siguiente:

Grafico 11.4.1 Tendencia

Teóricamente, si el proceso estocástico muestra decrecimiento o crecimiento significa que la serie no presenta un valor medio constante en todo el periodo muestral, es decir, no oscila en torno al mismo valor. En otros términos, una serie es estacionaria cuando su media y varianza permanecen constantes en el tiempo. Cuando se comprueba que una serie no presenta media y varianza constantes, entonces se dice que no es estacionaria. Por tanto, podemos suponer, a priori, que, probablemente, no es estacionaria, luego, presentará al menos una raíz unitaria (Pulido y López, 1999: 269, 273).

Comentario: Como puede observarse, la gráfica del proceso estocástico ARA presenta un claro comportamiento descendente o decreciente por lo que, podemos suponer que no es constante en media, luego entonces, podemos suponer que el proceso no es estacionario. Además, existen variables dentro del proceso que muestran gran dispersión, por lo que tampoco es constante en varianza (ver diagrama de dispersión en el cuadro siguiente). Derivado de lo cual, quizá sea necesario y/o preciso trabajar en diferencias. Además, tomando logaritmos se reduce la dispersión de la serie, es decir, la transformación en logaritmos la convierte en estacionaria en varianza. Sin embargo, realicemos otras pruebas gráficas y formales para determinarlo con precisión.

Grafico 11.4.2 Dispersión

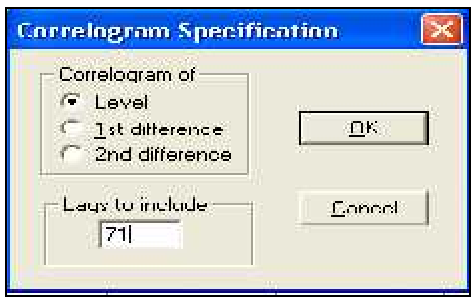


Nota: Para obtener el diagrama de dispersión de el proceso estocástico ARA vamos al menú principal y ahí seleccionamos: Quick/Graph/Scatter/, donde se desplegará una pantalla y en la cual escribiremos la serie a graficar, en nuestro caso ARA, y le damos Ok.

b) <u>Correlograma.</u>

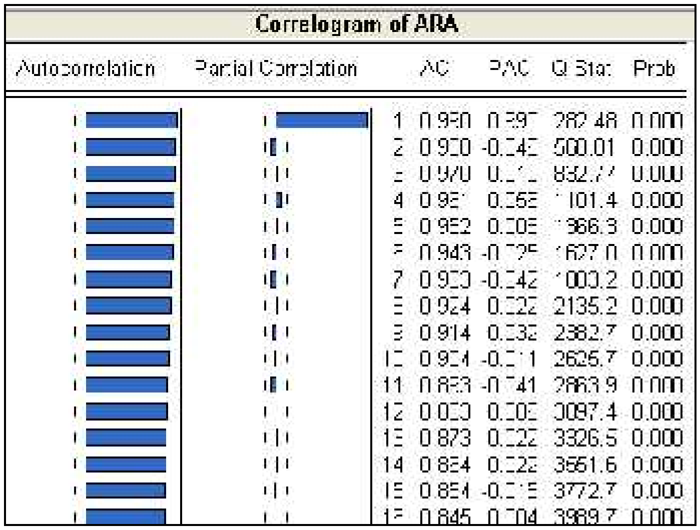
Otro procedimiento gráfico para comprobar la posible existencia de una raíz unitaria en el proceso estocástico consiste en inspeccionar el correlograma del mismo. Así, en la ventana de la serie vamos a: View/Correlogram…/, en donde se desplegará una ventana que nos solicita la especificación del correlograma a realizar, las opciones son: en niveles (Level), es decir, los valores de el proceso original, en primeras diferencias y en segundas diferencias. Asimismo, nos solicita en número de rezagos a incluir (Lags to include), mismo que por default Eviews da 36. La ventana es como la siguiente:

Imagen 11. 4.3 Comando



Sin embargo, siguiendo el criterio y/o sugerencia de Gujarati (2007: 784-785) en cuanto a el número de rezagos a utilizar y/o a la longitud del mismo, utilizaremos un cuarto de la longitud de la serie de tiempo, es decir, 71 variables aleatorias. Por lo que, para nuestro caso seleccionaremos el correlograma de niveles (Level) y en “Lags to include” escribiremos 71/Ok. Procedimiento que nos desplegará el correlograma siguiente:

Grafico 11.4.3 Correlograma



Correlogram of ARA

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.980	[illegible]	282.48	0.000
		2	0.9?0	-0.24?	560.01	0.000
		3	0.970	[illegible]	832.77	0.000
		4	0.96?	0.?5?	1101.4	0.000
		5	0.952	0.?0?	1366.3	0.000
		6	0.943	-0.?75	1627.0	0.000
		7	0.9?3	-0.?42	1?0?.2	0.000
		8	0.924	0.??2	2135.2	0.000
		9	0.914	0.?3?	2382.7	0.000
		10	0.904	-0.??1	2625.7	0.000
		11	0.893	-0.?41	2863.9	0.000
		12	0.8?3	0.?0?	3097.4	0.000
		13	0.873	0.??2	3326.5	0.000
		14	0.864	0.??2	3551.6	0.000
		15	0.854	-0.??5	3772.7	0.000
		16	0.845	0.?04	3989.7	0.000

En términos teóricos, si la función de autocorrelación muestral (primera columna de la izquierda del correlograma) decrece lentamente es muestra o indicio de la no estacionariedad de la serie, es decir, presenta una raíz unitaria. Asimismo, si la autocorrelación parcial (segunda columna de la izquierda del correlograma) muestra al menos un valor significativo en el retardo uno, con un coeficiente de autocorrelación parcial (PAC) cercano a la unidad (ρ ~ 1) es indicativo de la no estacionariedad de la serie (Pulido y López, 1999: 274).

Ahora bien, un correlograma que desciende rápidamente o es cuasialeatorio corresponde a variables estacionarias.

Asimismo, el estadístico "Q" también es una prueba de utilidad para determinar la no estacionariedad de la serie, pues si la probabilidad asociada a los rezagos es menor que 0.05 se dice que la serie es no estacionaria y, si la probabilidad asociada es mayor a 0.05 se concluye que la serie es estacionaria.

Sencillamente, cuando en el correlograma la función de autocorrelación desciende lentamente, así como si los valores de la autocorrelación muestral (AC) de los primeros rezagos son cercanos a 1, los primeros de la autocorrelación parcial (ACP) son cercanos a 1 y la probabilidad asociada a la estadística "Q" para cada rezago es menor a 0.05, no cabe duda de que se trata de una serie de tiempo no estacionaria.

Comentario: Dado que como puede observarse en el correlograma de la serie ARA la función de autocorrelación desciende lentamente y la función de autocorrelación parcial presenta un valor significativo en el primer coeficiente, cercano a la unidad, debemos proceder a la transformación de la serie en logaritmos y posteriormente a diferenciarla.

Asimismo, como todas las probabilidades asociadas al estadístico "Q" son menores a 0.05 se concluye que la serie es no estacionaria.

2. Método formal.

El método formal para analizar si una serie es estacionaria es el Test de raíces unitarias. Eviews incluye dos pruebas para probar la existencia de raíz unitaria, se trata de la prueba o Test de Dickey-Fuller (DF), una variante del mismo la Dickey-Fuller Aumentada (ADF) y el Test de Phillips-Perron (PP).

Para ambas pruebas se plantea la siguiente prueba de hipótesis:

Ho: $\rho = 1$; existe raíz unitaria, por tanto, la serie es no estacionaria.

Ha: $\rho < 1$; no existe raíz unitaria, por tanto, la serie es estacionaria.

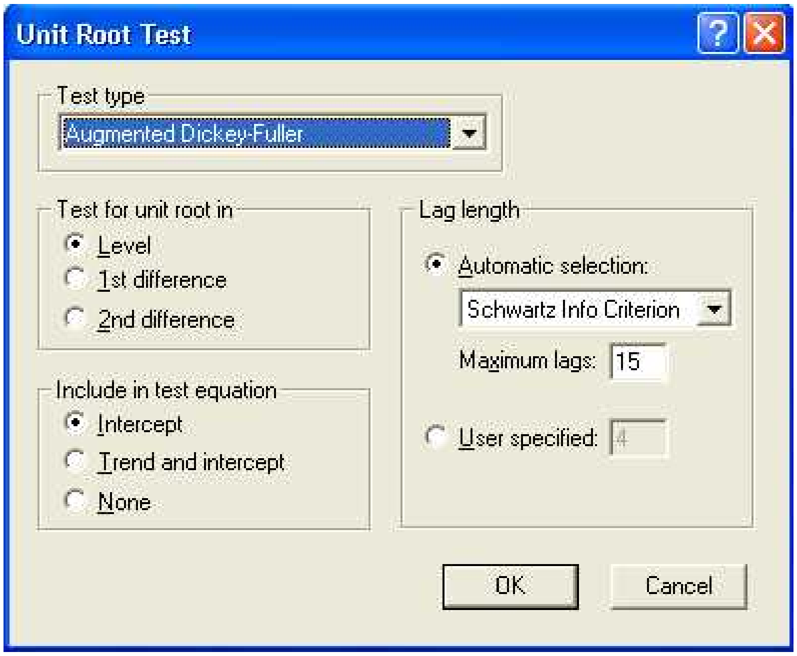
El criterio de aceptación de la Ho es que si el valor de t del estadístico de DF, ADF y PP es menor que los valores de t críticos de Mackinnon (a cualquier nivel de significación al que estemos trabajando), ambos en términos absolutos, se acepta la Ho. Si el valor de los estadísticos es mayor que los valores críticos de Mackinnon, ambos en términos absolutos, se acepta la Ha.

a) Test DF y ADF.

En Eviews para realizar la prueba DF y ADF una vez estando en la ventana de la serie vamos a: View/ Unit Root Test…/, en donde nos aparecerá una ventana como la siguiente. Cabe destacar que el Test de DF o simple es solamente válido si las series son un proceso autorregresivo de orden 1, AR(1).

El Test de ADF se utiliza cuando se tienen retardos o rezagos de orden superior al 1, por tanto, la serie sigue un proceso AR(ρ).

Imagen 11.4.4 Prueba Unit Root Test



En primer lugar, hay que especificar el número de retardos de los términos de las primeras diferencias de la serie para incluir en la regresión del Test. En Eviews tenemos dos opciones para seleccionar el número de retardos (Lag length): De forma automática (Automatic selection) y especificado por el usuario (User specified). Si, por ejemplo, seleccionamos el especificado por el usuario e incluimos cero (0) retardos o rezagos, Eviews nos proporciona el Test de DF y si indicamos un número mayor que cero, generará el Test de ADF.

En segundo lugar, el programa pregunta acerca de la inclusión de otras variables exógenas en la regresión auxiliar del Test. Podemos incluir una constante (Intercept), una constante y un término de tendencia lineal (Trend and intercept) o, simplemente, no incluir nada adicional en la regresión del Test (None). ¿Qué variables incluir entonces? Podemos aplicar las siguientes normas de carácter general:

i) Si la serie original presenta tendencia (creciente o decreciente), se deberán incluir ambos términos en la regresión, es decir, considerar como regresores al término independiente y al término de tendencia lineal;

ii) Si la serie original no parece presentar tendencia y tiene un valor medio distinto de cero debemos incluir un término constante en la regresión; y

iii) Si la serie original parece fluctuar en torno al valor medio cero, no se considera necesario incluir ningún regresor adicional en la regresión, es decir, no incluimos ni constante ni término de tendencia (Pulido y López, 1999: 271).

Entonces, una vez estando en la ventana de la serie, como se comento previamente, vamos a: View/Unit Root Test…/, con lo que aparecerá una ventana de dialogo como la anterior, en la cual seleccionaremos el Test de ADF teniendo siempre presente las consideraciones anteriores. Para aplicar el Test de DF:

- Seleccionamos “Augmented Dickey-Fuller” en la especificación del Test (Test Type);

- Después, indicamos si el Test se va a realizar sobre la serie en niveles u original (Level), en primeras diferencias o en segundas diferencias de la serie original en “Test for unit root in”;

- Especificamos si queremos incluir una constante (intercept), una constante más un término de tendencia lineal (Trend and intercept) o ninguno de los dos casos anteriores (None); y

- Finalmente, debemos especificar el orden de la correlación serial a considerar en las series. Para el Test de ADF, se trata de especificar el número de retardos de los términos de primeras diferencias de la serie a añadir en la regresión del Test (Pulido y López, 1999: 272).

Algunas consideraciones generales sobre la selección de los datos es que es muy importante y hay que considerar lo siguiente:

i) Si el Test acepta la Ho en los datos en niveles, pero rechaza la Ho en primeras diferencias, entonces la serie contiene una raíz unitaria y, por tanto, es integrada de orden 1, I(1); y

ii) Si el Test acepta la Ho en niveles y primeras diferencias, pero se rechaza al realizar el Test en segundas diferencias, entonces la serie contiene dos raíces unitarias y es integrada de orden 2, I(2).

Para nuestro caso, vamos a realizar los dos Test, es decir, el DF y el ADF.

➢ **Test DF.**

Dado que la serie original de ARA presenta tendencia decreciente vamos a incluir en el Test el intercepto y el término de tendencia lineal; luego, vamos a seleccionar los rezagos de forma manual o especificados por el usuario, en donde indicaremos que son cero para trabajar con el Test de DF, y; finalmente, vamos a realizar el Test primero en niveles, luego en primeras diferencias para aceptar o rechazar la Ho. Así, siguiendo los pasos preestablecidos, el resultado para el Test de DF en niveles es:

Cuadro 11.4.1 Prueba Dickey Fuller Aumentada

Null Hypothesis: ARA has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-2.438438**	**0.3589**
Test critical values:	**1% level**	**-3.990701**	
	5% level	**-3.425728**	
	10% level	**-3.136027**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ARA)
Method: Least Squares
Date: 03/01/09 Time: 23:50
Sample(adjusted): 2 285
Included observations: 284 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ARA(-1)	-0.036564	0.014995	-2.438438	0.0154
C	0.466007	0.199219	2.339171	0.0200
@TREND(1)	-0.001323	0.000536	-2.470401	0.0141

R-squared	0.021590	Mean dependent var	-0.027254
Adjusted R-squared	0.014627	S.D. dependent var	0.241696
S.E. of regression	0.239922	Akaike info criterion	-0.006496
Sum squared resid	16.17512	Schwarz criterion	0.032050
Log likelihood	3.922429	F-statistic	3.100400
Durbin-Watson stat	1.890580	Prob(F-statistic)	0.046575

Comentario: Como el valor de de t de DF, 2.43, es menor que el valor t crítico de MacKinnon al 5%, en términos absolutos, de significación estadística, 3.42, se acepta la Ho. Luego, entonces, existe raíz unitaria, por tanto, la serie es no estacionaria. Asimismo, al ser la probabilidad asociada al estadístico DF mayor que 0.05 se acepta la Ho y se rechaza la Ha.

Ahora, realicemos el Test de DF pero en primeras diferencias con los demás criterios constantes, así obtenemos:

Cuadro 11.4.2 Prueba Dickey Fuller Aumentada

Null Hypothesis: D(ARA) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-16.12400**	**0.0000**
Test critical values:	**1% level**	**-3.990817**	
	5% level	**-3.425784**	
	10% level	**-3.136061**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ARA,2)
Method: Least Squares
Date: 03/01/09 Time: 23:57
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ARA(-1))	-0.962111	0.059669	-16.12400	0.0000
C	-0.017313	0.029055	-0.595879	0.5517
@TREND(1)	-6.78E-05	0.000176	-0.384186	0.7011

R-squared	0.481487	Mean dependent var	-0.000353
Adjusted R-squared	0.477784	S.D. dependent var	0.335380
S.E. of regression	0.242361	Akaike info criterion	0.013768
Sum squared resid	16.44689	Schwarz criterion	0.052412
Log likelihood	1.051834	F-statistic	130.0031
Durbin-Watson stat	1.984524	Prob(F-statistic)	0.000000

Comentario: Dado que en términos absolutos el valor del estadístico DF, 16.12, es mayor que el valor crítico de MacKinnon al 5% de significación igual a 3.42, se rechaza la Ho y se acepta la Ha, por tanto, no existe raíz unitaria, es decir, la serie es estacionaria. Asimismo, al ser la probabilidad asociada al estadístico DF menor a 0.05 se acepta la Ha y se rechaza la Ho.

Conclusión preliminar: En virtud de que el Test de DF en primeras diferencias rechaza la Ho y acepta la Ha y dado que en niveles se acepta la Ho y se rechaza la Ha concluimos que la serie contiene una raíz unitaria y, por tanto, es integrada de orden 1, I(1).

- **Test ADF.**

Para elaborar este Test vamos a modificar únicamente el número de rezagos, así, en lugar de incluir cero como en el Test anterior de DF, vamos a incluir 1. Los demás términos los mantendremos constantes, de tal suerte que como la serie original de ARA presenta tendencia decreciente vamos a incluir en el Test el intercepto y el término de tendencia lineal; luego, vamos a seleccionar los rezagos de forma manual o especificados por el usuario, en donde indicaremos que es uno para trabajar con el Test de ADF, y; finalmente, vamos a realizar el Test primero en niveles y luego en primeras diferencias para aceptar o rechazar la Ho. Así, siguiendo los pasos descritos, el resultado para el Test de ADF en niveles es:

Cuadro 11.4.3 Prueba Dickey Fuller Aumentada

Null Hypothesis: ARA has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-2.445320**	**0.3554**
Test critical values:	**1% level**	**-3.990817**	
	5% level	**-3.425784**	
	10% level	**-3.136061**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ARA)
Method: Least Squares
Date: 03/02/09 Time: 01:21
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ARA(-1)	-0.037107	0.015175	-2.445320	0.0151
D(ARA(-1))	0.053875	0.059506	0.905367	0.3661
C	0.471624	0.202012	2.334639	0.0203
@TREND(1)	-0.001325	0.000543	-2.439810	0.0153

R-squared	0.022963	Mean dependent var	-0.028057
Adjusted R-squared	0.012457	S.D. dependent var	0.241745
S.E. of regression	0.240234	Akaike info criterion	-0.000371
Sum squared resid	16.10179	Schwarz criterion	0.051155
Log likelihood	4.052452	F-statistic	2.185769
Durbin-Watson stat	1.983966	Prob(F-statistic)	0.089938

Comentario: Dado que el valor del estadístico de ADF es menor en términos absolutos al valor crítico de MacKinnon, a saber 2.44 < 3.42 al 5% de significación estadística, y en virtud de que la probabilidad asociada al estadístico ADF es mayor a 0.05 se acepta la Ho y por tanto se dice que existe raíz unitaria, es decir, la serie es no estacionaria.

El resultado del Test de ADF con primeras diferencias manteniendo constante los demás términos es:

Cuadro 11.4.4 Prueba Dickey Fuller Aumentada

Null Hypothesis: D(ARA) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-11.88449**	**0.0000**
Test critical values:	**1% level**	**-3.990935**	
	5% level	**-3.425841**	
	10% level	**-3.136094**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ARA,2)
Method: Least Squares
Date: 03/02/09 Time: 01:36
Sample(adjusted): 4 285
Included observations: 282 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ARA(-1))	-0.986810	0.083033	-11.88449	0.0000
D(ARA(-1),2)	0.029472	0.059800	0.492843	0.6225
C	-0.013561	0.029296	-0.462898	0.6438
@TREND(1)	-9.19E-05	0.000178	-0.517239	0.6054

R-squared	0.480514	Mean dependent var	0.001383
Adjusted R-squared	0.474908	S.D. dependent var	0.334700
S.E. of regression	0.242534	Akaike info criterion	0.018737
Sum squared resid	16.35279	Schwarz criterion	0.070395
Log likelihood	1.358073	F-statistic	85.71467
Durbin-Watson stat	2.008458	Prob(F-statistic)	0.000000

Comentario: Como puede observarse, al ser el valor del Test ADF mayor que el valor crítico de MacKinnon (11.88 > 3.42) en términos absolutos al 5% de significación estadística aceptamos la Ha; además, al ser la probabilidad asociada al Test de ADF menor a 0.05 se corrobora la aceptación de la Ha, y se concluye que no existe raíz unitaria, por tanto, la serie es estacionaria.

Conclusión de las Pruebas DF y ADF: Derivado de los resultados obtenidos y en virtud de que con ambas pruebas se acepta la Ho con los datos en niveles, pero se rechaza la Ho en primeras diferencias, entonces la serie contiene una raíz unitaria y, por tanto, es no estacionaria y es integrada de orden 1, I(1). Por lo que, de acuerdo a los resultados obtenidos, fue necesario diferenciar una vez la serie para volverla estacionaria.

b) Test PP.

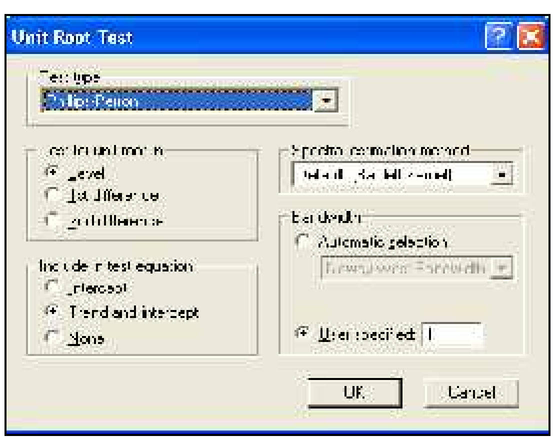
Este Test es similar a la prueba de ADF, pues si el valor del estadístico PP es menor o inferior a los valores críticos de MacKinnon, a cualquier nivel de significación en términos absolutos en ambos casos, se acepta la Ho de existencia de una raíz unitaria.

Mientras que el Test de ADF corrige la correlación serial de orden elevado añadiendo más retardos del término diferenciado de la serie original en el lado derecho de la ecuación, el Test de PP realiza una corrección del estadístico “τ” sobre el coeficiente en la regresión AR(1) para considerar la correlación serial en el término “e”.

En este Test, al igual que en el ADF, tenemos que especificar si incluimos una constante, constante más término de tendencia lineal o nada en la regresión. Para este Test hay que especificar el número de periodos de correlación serial a incluir, es decir, rezagos o retardos Pulido y López, 1999: 271).

En Eviews estando en la ventana de la serie vamos a: View/Unit Root Test… (test de raíces unitarias)/, en donde seleccionaremos el Test de Phillips-Perron (PP), en “test for unit root in” seleccionamos primeramente niveles (Level), en “Include in test equation” seleccionamos “Trend and intercept”, en “Spectral estimation method” vamos a dejar el que por default nos proporciona Eviews, es decir, el Bartlett Kernel, y en “Bandwidth” vamos a seleccionar especificado por el usuario (User specified) en donde escribiremos 1 y Ok. De tal suerte que, la ventana de raíces unitarias con el Test de PP debe visualizarse de la siguiente manera:

Imagen 11.4.5 Comando



El resultado del Test de PP en niveles, con intercepto y término de tendencia lineal y bandwidth especificado por el usuario de 1 es:

Cuadro 11.4.5 Prueba Phillips Perron

Null Hypothesis: ARA has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		**-2.485894**	**0.3350**
Test critical values:	**1% level**	**-3.990701**	
	5% level	**-3.425728**	
	10% level	**-3.136027**	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.056955
HAC corrected variance (Bartlett kernel)	0.059988

Phillips-Perron Test Equation
Dependent Variable: D(ARA)
Method: Least Squares
Date: 03/02/09 Time: 16:54
Sample(adjusted): 2 285
Included observations: 284 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ARA(-1)	-0.036564	0.014995	-2.438438	0.0154
C	0.466007	0.199219	2.339171	0.0200
@TREND(1)	-0.001323	0.000536	-2.470401	0.0141

R-squared	0.021590	Mean dependent var	-0.027254
Adjusted R-squared	0.014627	S.D. dependent var	0.241696
S.E. of regression	0.239922	Akaike info criterion	-0.006496
Sum squared resid	16.17512	Schwarz criterion	0.032050
Log likelihood	3.922429	F-statistic	3.100400
Durbin-Watson stat	1.890580	Prob(F-statistic)	0.046575

Comentario: Como el valor del estadístico PP es, es términos absolutos, menor que el valor crítico de MacKinnon al 5% de significación estadística (2.48 < 3.42) se acepta la Ho y dado que la probabilidad asociada al estadístico PP es mayor a 0.05 se acepta la Ho de que existe raíz unitaria, por tanto, la serie es no estacionaria.

Ahora, realicemos la prueba PP en primeras diferencias manteniendo constante los demás términos de la selección, así obtenemos:

Cuadro 11.4.6 Prueba Phillips Perron

Null Hypothesis: D(ARA) has a unit root			
Exogenous: Constant, Linear Trend			
Bandwidth: 1 (Fixed using Bartlett kernel)			
		Adj. t-Stat	**Prob.***
Phillips-Perron test statistic		**-16.12566**	**0.0000**
Test critical values:	**1% level**	**-3.990817**	
	5% level	**-3.425784**	
	10% level	**-3.136061**	
*MacKinnon (1996) one-sided p-values.			

Residual variance (no correction)	0.058116
HAC corrected variance (Bartlett kernel)	0.058396

Phillips-Perron Test Equation
Dependent Variable: D(ARA,2)
Method: Least Squares
Date: 03/02/09 Time: 16:59
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ARA(-1))	-0.962111	0.059669	-16.12400	0.0000
C	-0.017313	0.029055	-0.595879	0.5517
@TREND(1)	-6.78E-05	0.000176	-0.384186	0.7011

R-squared	0.481487	Mean dependent var	-0.000353
Adjusted R-squared	0.477784	S.D. dependent var	0.335380
S.E. of regression	0.242361	Akaike info criterion	0.013768
Sum squared resid	16.44689	Schwarz criterion	0.052412
Log likelihood	1.051834	F-statistic	130.0031
Durbin-Watson stat	1.984524	Prob(F-statistic)	0.000000

Comentario: Dado que el valor del estadístico PP es mayor, en términos absolutos, que el valor crítico de MacKinnon al 5% (16.12 > 3.42) y en virtud de que la probabilidad asociada al estadístico es menor a 0.05, aceptamos la Ha y rechazamos la Ho, por tanto, no existe raíz unitaria, es decir, la serie es estacionaria.

Conclusión del Test PP: Dado que con la prueba PP en niveles se acepta la Ho y en primeras diferencias se acepta la Ha, concluimos que la serie original "ARA" no es estacionaria y por tanto hay que diferenciarla una vez para convertirla en estacionaria.

Así, tanto los métodos informales como los formales nos indican que la serie original "ARA" es no estacionaria, por lo que es necesario diferenciarla una vez para volverla estacionaria. Luego entonces, la serie es integrada de orden 1, I(1). Asimismo, es preciso, antes de diferenciarla, convertirla en logaritmos en virtud de la dispersión que presentan algunas variables aleatorias, lo cual es indicio de la inexistencia de varianza constante.

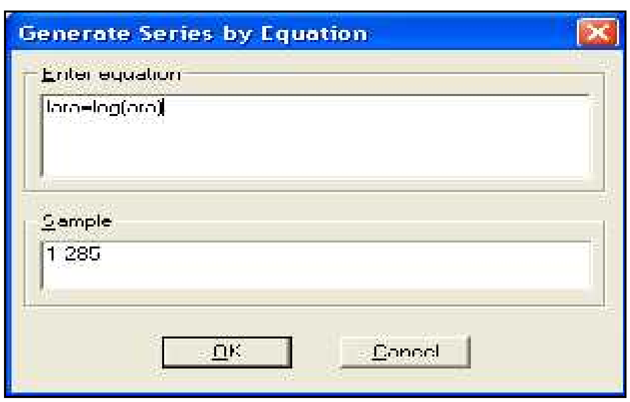
II) <u>Transformación y diferenciación de la serie.</u>

Teóricamente, para obtener una <u>serie en media constante</u> es preciso obtener sus primeras diferencias y, para obtener <u>una serie con varianza constante</u> es necesario transformarla en logaritmos antes de diferenciarla, ya que las series transformadas en logaritmos presentan menor dispersión.

Recuérdese que el orden de integración "I" <u>es el número de operaciones de diferenciación que hay que efectuar para convertir la serie en estacionaria</u> (Pulido y López, 1999: 269).

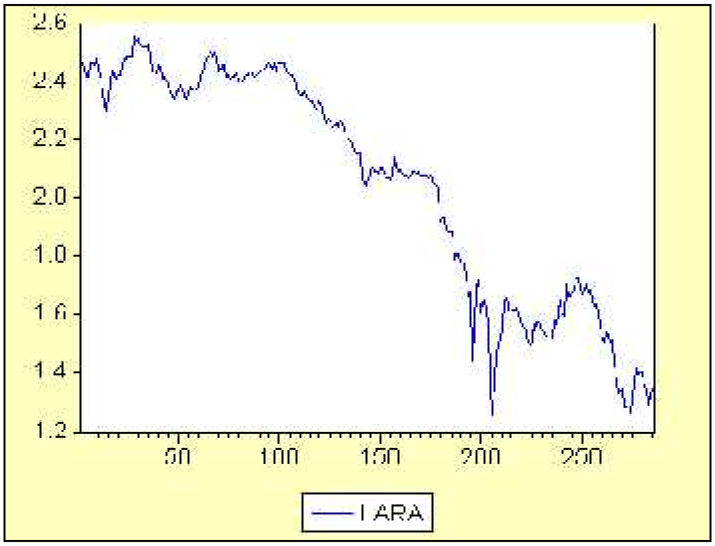
De esta forma, primero debemos generar la serie en logaritmos y graficarla de forma lineal y a través del diagrama de dispersión para ver si disminuyó la dispersión que presenta. Así, vamos al menú principal y seleccionamos: Quick/Generate Series…/, y en la ventana que nos desplegará escribiremos: lara=log(ara)/Ok., donde "l" indica que la serie fue transformada en logaritmos, ello se observa de la siguiente manera:

Imagen 11.4.6 Comando



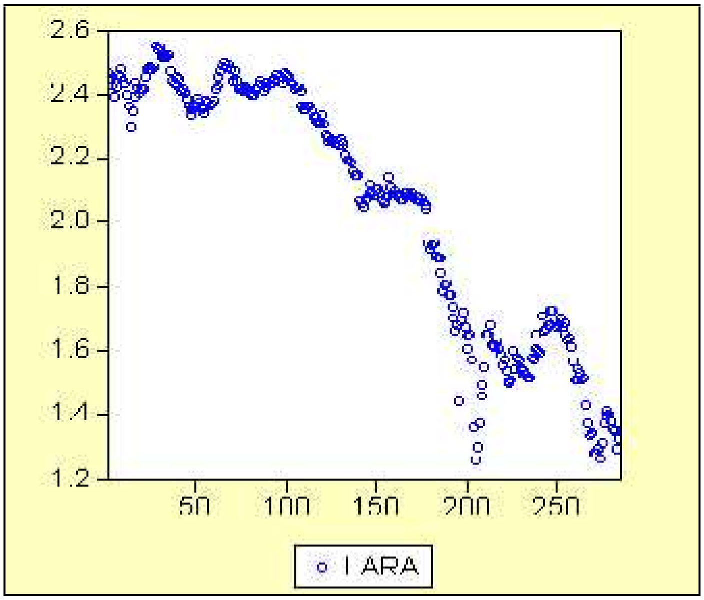
La nueva variable generada (lara) aparecerá en la ventana del Workfile. A continuación procedemos a graficarla de forma lineal a través de los pasos ya descritos anteriormente. De esta forma obtenemos:

Grafico 11.4.4 Tendencia



Su diagrama de dispersión es:

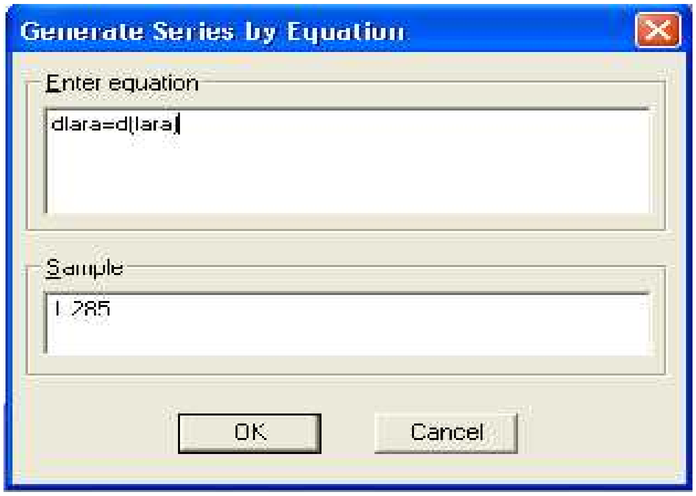
Grafico 11.4.5 Dispersión



Como puede observarse, la transformación en logaritmos de la serie (lara) aun presenta un comportamiento similar al de la serie original (ara), es decir, sin transformar. Diferenciemos entonces la serie transformada en logaritmos una vez y grafiquémosla para ver su comportamiento.

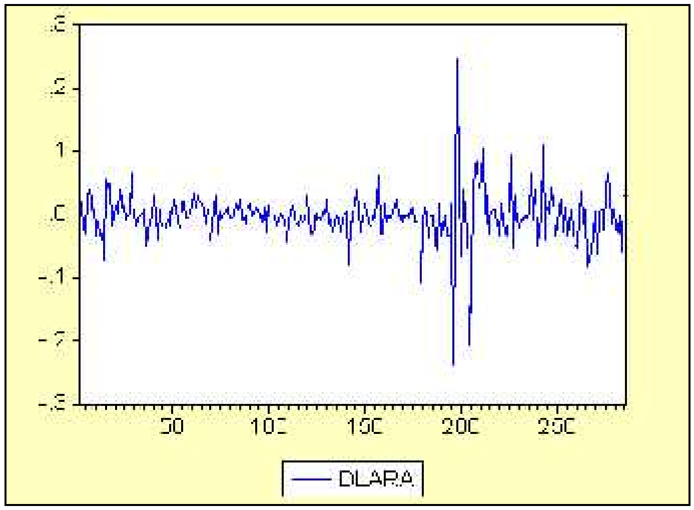
El procedimiento para diferenciar la nueva serie en logaritmos (lara) es el siguiente, en el menú principal vamos a: Quick/Generate Series…/, en la ventana que desplegará escribiremos dlara=d(lara)/Ok., donde "d" indica que se trata de la primera diferencia de la serie "lara". Ello se observa así:

Imagen 11.4.7 Comando



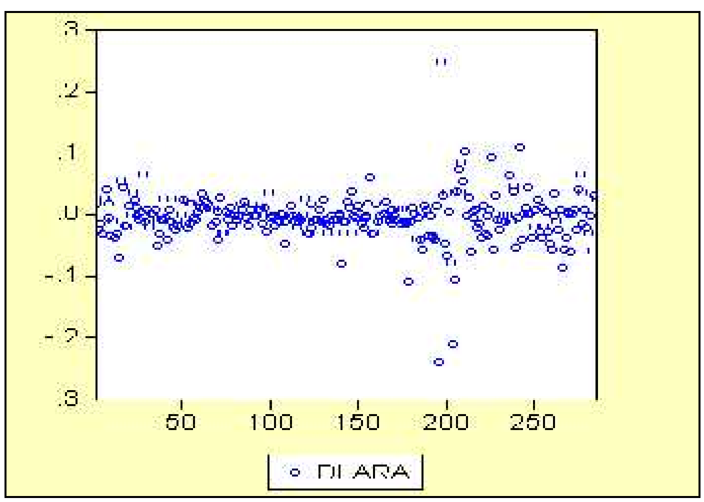
La nueva serie en logaritmos y diferenciada deberá aparecer en la ventana del Workfile. Una vez ahí, vamos a abrir la serie "dlara" dando doble click sobre la misma y a continuación vamos a graficarla a través del procedimiento conocido, es decir, en la ventana de la serie vamos a: View/Graph/Line/Ok., lo que nos desplegará la gráfica siguiente:

Grafico 11.4.6 Tendencia



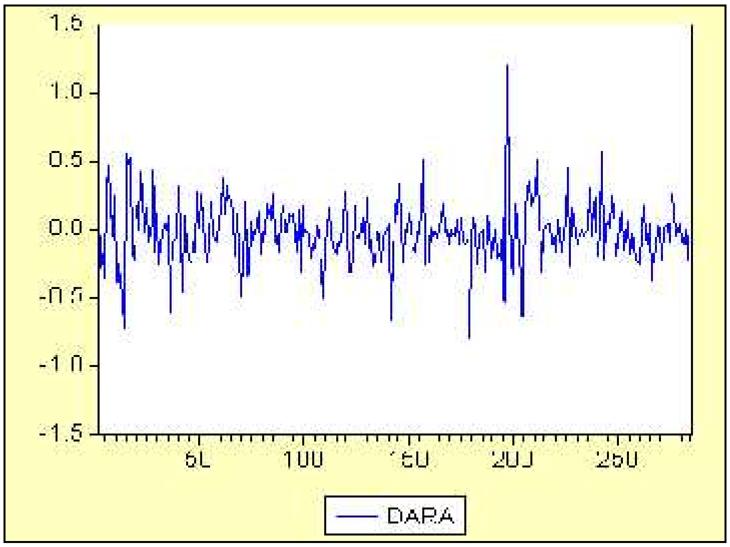
Como puede apreciarse, la serie “dlara” presenta gráficamente un comportamiento paralelo al eje de las abscisas, lo que indica una media constante y la no existencia de tendencia. Asimismo, el diagrama de dispersión nos muestra que la mayoría de las variables aleatorias oscilan alrededor de la media como puede observarse en el diagrama de dispersión de “dlara”.

Grafico 11.4.8 Tendencia



Pero, ¿Será adecuada la transformación y diferenciación correspondiente que realizamos? La forma de saberlo es crear directamente la serie “ara” en diferencias y obtener su gráfico correspondiente. Para ello, vamos al menú principal y ahí seleccionamos: Quick/Generate Series…/, y en la ventana que nos desplegará escribimos dara=d(ara). De la cual su gráfico correspondiente es:

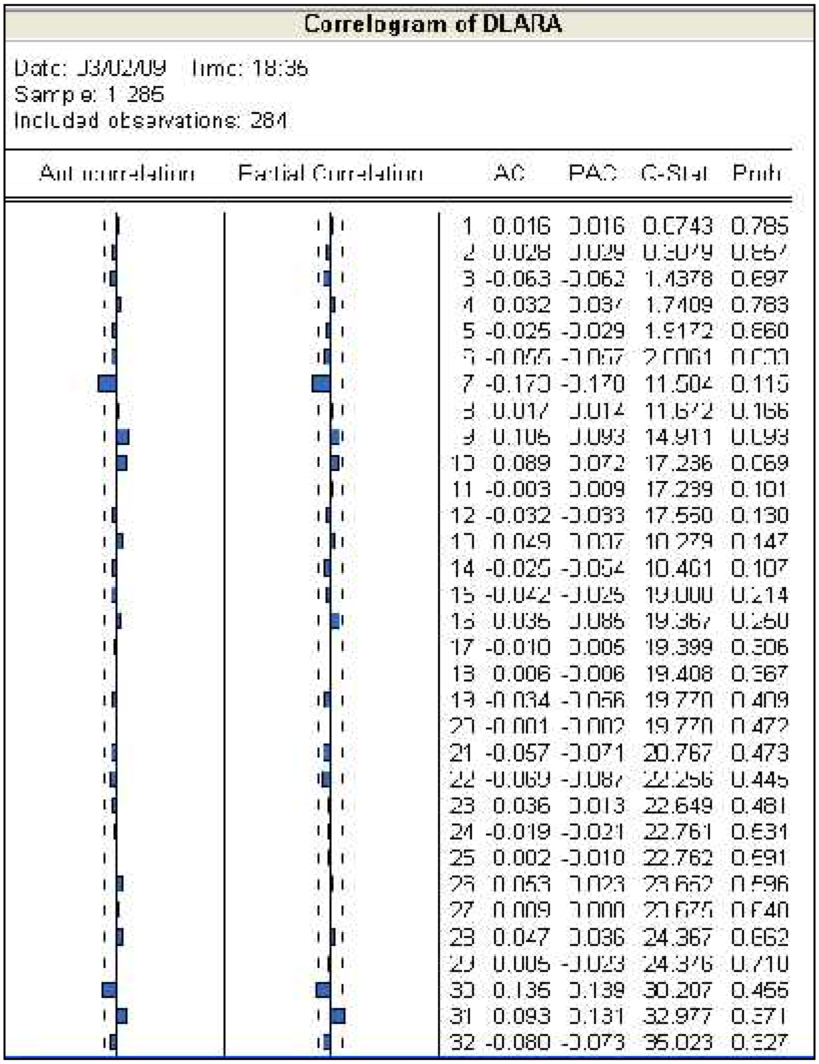
Grafico 11.4.9 Tendencia



Como puede observarse, en la gráfica de la serie “dlara” los valores de las variables aleatorias oscilan en torno al 0.2 y al -0.2, mientras que en la serie “dara” dichos valores oscilan en torno al 1.0 y al -1.0, por tanto, los valores de la serie transformada en logaritmos y diferenciada una vez presentan menor dispersión, lo que supone una media y varianza constante. Entonces, para nuestro caso no cabe duda de la serie con la que vamos a trabajar, es decir, aquella que fue transformada en logaritmos y diferenciada una vez (dlara) pues, es estacionaria.

Con ello, podemos ahora realizar el correlograma de la serie “dlara” para verificar la estacionariedad de la misma. Así, con los pasos ya conocidos obtenemos el siguiente correlograma:

Grafico 11.4.10 Correlograma



Correlogram of DLARA

Date: 03/02/09 Time: 18:36
Sample: 1 285
Included observations: 284

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.016	0.016	0.0743	0.785
		2	0.028	0.029	0.3079	0.857
		3	-0.063	-0.062	1.4378	0.697
		4	0.032	0.037	1.7409	0.783
		5	-0.025	-0.029	1.9172	0.860
		6	-0.055	-0.057	2.8061	0.833
		7	-0.170	-0.170	11.504	0.118
		8	0.017	0.014	11.672	0.166
		9	0.106	0.093	14.911	0.093
		10	0.089	0.072	17.236	0.069
		11	-0.003	0.009	17.239	0.101
		12	-0.032	-0.033	17.550	0.130
		13	0.049	0.037	18.279	0.147
		14	-0.025	-0.054	18.461	0.187
		15	-0.042	-0.025	19.000	0.214
		16	0.036	0.086	19.367	0.250
		17	-0.010	0.005	19.399	0.306
		18	0.006	-0.006	19.408	0.367
		19	-0.034	-0.056	19.770	0.409
		20	-0.001	-0.002	19.770	0.472
		21	-0.057	-0.071	20.767	0.473
		22	-0.069	-0.087	22.256	0.445
		23	0.036	0.013	22.649	0.481
		24	-0.019	-0.021	22.761	0.531
		25	0.002	-0.010	22.762	0.591
		26	0.053	0.023	23.652	0.596
		27	0.009	0.000	23.675	0.648
		28	0.047	0.036	24.367	0.662
		29	0.005	-0.023	24.376	0.710
		30	0.136	0.139	30.207	0.455
		31	0.093	0.131	32.977	0.371
		32	-0.080	-0.073	35.023	0.327

Como puede observarse, al encontrarse los valores de la autocorrelación (AC) y los de la autocorrelación parcial (PAC) dentro de las bandas de confianza, al ser los valores de AC y PAC muy menores a 1 y al ser la probabilidad asociada al estadístico “Q” mayores a 0.05 concluimos, de acuerdo a este correlograma, que la serie “dlara” es estacionaria.

Ahora, realicemos los Test de raíces unitarias correspondientes, es decir, las pruebas o métodos formales, para determinar que la serie efectivamente es ya estacionaria, pues, gráficamente lo es.

a) **Test de DF para "dlara".**

Siguiendo los pasos ya descritos con antelación para realizar la prueba DF y, planteando la prueba de hipótesis correspondiente de la siguiente manera:

Ho: $\rho = 1$; existe raíz unitaria, por tanto, la serie es no estacionaria.

Ha: $\rho < 1$; no existe raíz unitaria, por tanto, la serie es estacionaria.

Cabe hacer mención que para realizar la prueba DF va a hacer en niveles (Level) porque ya está diferenciada una vez, por lo que si se selecciona primeras diferencias en realidad se tratará de las segundas diferencias; dado que la nueva serie (dlara) no parece presentar tendencia (creciente o decreciente) pero parece tener uno o más valores medios distintos de cero debemos incluir un término constante (intercept) en la regresión; y en el número de rezagos cero para realizar el Test DF. El resultado del Test de DF es:

Cuadro 11.4.7 Prueba Dickey Fuller Aumentada

Null Hypothesis: DLARA has a unit root
Exogenous: Constant
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-16.48547**	**0.0000**
Test critical values:	**1% level**	**-3.453317**	
	5% level	**-2.871546**	
	10% level	**-2.572174**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DLARA)
Method: Least Squares
Date: 03/02/09 Time: 18:58
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLARA(-1)	-0.983876	0.059681	-16.48547	0.0000
C	-0.003884	0.002284	-1.700752	0.0901

R-squared	0.491652	Mean dependent var	3.23E-05
Adjusted R-squared	0.489843	S.D. dependent var	0.053500
S.E. of regression	0.038213	Akaike info criterion	-3.684263
Sum squared resid	0.410316	Schwarz criterion	-3.658500
Log likelihood	523.3232	F-statistic	271.7706
Durbin-Watson stat	1.993651	Prob(F-statistic)	0.000000

Como el valor del estadístico DF (16.48) en términos absolutos es mayor que el valor crítico de MacKinnon al 5% (2.87) y dado que la probabilidad del estadístico DF es menor a 0.05 se acepta la Ha y se rechaza la Ho, por tanto, no existe raíz unitaria, es decir, la serie "dlara" es estacionaria.

b) Test de ADF para "dlara".

Ahora realicemos la prueba ADF, manteniendo los términos seleccionados constantes pero modificando el número de rezagos a 1 en lugar de cero. Obtenemos:

Cuadro 11.4.8 Prueba Dickey Fuller Aumentada

Null Hypothesis: DLARA has a unit root
Exogenous: Constant
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-12.01737**	**0.0000**
Test critical values:	**1% level**	**-3.453400**	
	5% level	**-2.871582**	
	10% level	**-2.572193**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DLARA)
Method: Least Squares
Date: 03/02/09 Time: 19:03
Sample(adjusted): 4 285
Included observations: 282 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLARA(-1)	-1.011255	0.084149	-12.01737	0.0000
D(DLARA(-1))	0.028867	0.059940	0.481598	0.6305
C	-0.003922	0.002307	-1.699698	0.0903

R-squared	0.491521	Mean dependent var	0.000181
Adjusted R-squared	0.487876	S.D. dependent var	0.053536
S.E. of regression	0.038312	Akaike info criterion	-3.675520
Sum squared resid	0.409521	Schwarz criterion	-3.636777
Log likelihood	521.2484	F-statistic	134.8477
Durbin-Watson stat	2.002513	Prob(F-statistic)	0.000000

En virtud de que el valor del estadístico ADF (12.01) es mayor que el valor crítico de MacKinnon al 5% (2.87) y dado que la probabilidad asociada al estadístico ADF es menor a 0.05 se acepta la Ha, por tanto, no existe raíz unitaria, es decir, la serie "dlara" es estacionaria.

c) **Test de PP para "dlara".**

Para el caso del Test de PP vamos a realizarlo en niveles (Level); con intercepto (intercept); y vamos a especificar de forma manual 1 rezago. Así obtenemos lo siguiente:

Cuadro 11.4.9 Prueba Phillips Perron

Null Hypothesis: DLARA has a unit root
Exogenous: Constant
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		**-16.48574**	**0.0000**
Test critical values:	**1% level**	**-3.453317**	
	5% level	**-2.871546**	
	10% level	**-2.572174**	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.001450
HAC corrected variance (Bartlett kernel)	0.001452

Phillips-Perron Test Equation
Dependent Variable: D(DLARA)
Method: Least Squares
Date: 03/02/09 Time: 19:11
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLARA(-1)	-0.983876	0.059681	-16.48547	0.0000
C	-0.003884	0.002284	-1.700752	0.0901

R-squared	0.491652	Mean dependent var	3.23E-05
Adjusted R-squared	0.489843	S.D. dependent var	0.053500
S.E. of regression	0.038213	Akaike info criterion	-3.684263
Sum squared resid	0.410316	Schwarz criterion	-3.658500
Log likelihood	523.3232	F-statistic	271.7706
Durbin-Watson stat	1.993651	Prob(F-statistic)	0.000000

Como puede observarse en los resultados del Test de PP el estadístico PP es mayor (16.48) al valor crítico de MacKinnon al 5% (2.87) y dado que la probabilidad asociada al estadístico PP es menor a 0.05, entonces aceptamos la Ha y rechazamos la Ho, por lo que no existe raíz unitaria, luego entonces la serie “dlara” es estacionaria.

***Conclusión general*:** Dado que con los métodos informales (gráfica lineal, diagrama de dispersión y correlograma) concluimos que la serie “dlara”, misma que fue transformada en logaritmos y diferenciada una vez, es estacionaria; y con los métodos formales (Test de DF, ADF y PP) se acepta la Ha en virtud de que los valores de los estadísticos DF, ADF y PP son mayores a los valores críticos de MacKinnon al 5% y a que la probabilidad asociada a los mismos es menor a 0.05 concluimos contundentemente que la serie de tiempo “dlara” es estacionaria. Por tanto, podemos pasar a la identificación del modelo ARMA, pues ya sabemos que la serie es integrada de orden 1, es decir, I(1).

III) Identificación de la estructura ARMA.

Una vez que hemos transformado la serie en estacionaria o en su caso probado que la serie es estacionaria de origen procedemos a determinar el tipo de modelo más adecuado para la serie objeto de estudio, es decir, el orden de los procesos autorregresivos (AR) y de medias móviles (MA), esto es, los valores de p y q.

Sin embargo, no se piense que resulta fácil, incluso para personas con experiencia, seleccionar el modelo ARMA. En muchas ocasiones se tendrán dudas razonables sobre el modelo más adecuado. Al respecto, Pulido y López (1999: 294, 301) sostienen que “no debemos preocuparnos en exceso por ello, ya que una posibilidad es seleccionar varios modelos alternativos y comprobar posteriormente cuál de ellos resulta finalmente más idóneo. Generalmente, suele procederse seleccionando siempre procesos lo más sencillos posible”. Asimismo, Johnston y Dinardo (2001) reconocen lo difícil que resulta la elección del orden de p y q. No obstante, sostienen que esta elección se agiliza mediante un procedimiento numérico sugerido por Hannan y Rissanen. Se trata de un procedimiento en tres etapas, a saber:

1.- Se estiman por MCO procesos AR puros de orden bastante elevado, lo cual no deja de ser razonable ya que un proceso ARMA desconocido equivale a un proceso AR infinito;

2.- Se selecciona la regresión que ha proporcionado el menor valor según el criterio de información de Akaike (AIC) y los residuos de dicha regresión se consideran como estimadores de los errores de un modelo ARMA; y

3.- Se ajustan algunos modelos ARMA utilizando para ello los residuos estimados. Después del ajuste, es decir, de la inclusión de MA(q) elegimos como especificación correcta aquella que proporciona el menor valor del criterio de Schwarz.

Técnicamente, la decisión de qué proceso a incluir en el modelo y su orden se tomará a partir de las funciones de autocorrelación (AC) y autocorrelación parcial (PAC). Es decir, el instrumento técnico básico para identificar un modelo ARMA es la denominada función de autocorrelación, misma que mide el grado de correlación entre cada valor de la variable y los desfasados 1, 2, …, h periodos (Pulido y López, 1999: 285-286). Habitualmente se terminará eligiendo entre los procesos más simples: AR(1), AR(2), MA(1), MA(2) y ARMA(1,1). En caso de duda pueden seleccionarse varios modelos alternativos, que serán estimados y contrastados posteriormente, para decidir el definitivamente adoptado.

Por su parte, para identificar un proceso AR, si la función de autocorrelación decrece (en forma regular, es decir, alternando valores positivos y negativos o en forma de ondas sinusoidales) y la función de autocorrelación parcial presenta un número de coeficientes igual al orden del proceso, entonces, el AR corresponderá a el número de coeficientes de la autocorrelación parcial que sobrepasan las bandas de confianza.

Un modelo MA(1) tiene sólo el primer coeficiente distinto de cero en la función de autocorrelación y la función de autocorrelación parcial muestra un comportamiento decreciente; un modelo MA(2) tendrá dos coeficientes no nulos en la función de autocorrelación mientras que la función de autocorrelación parcial mostrará un comportamiento decreciente.

Un modelo ARMA(1,1) presentará diversas combinaciones, pero siempre mostrará un comportamiento decreciente en ambas funciones, es decir, la de autocorrelación y la de autocorrelación parcial.

Por tanto, a la hora de identificar un modelo ARMA debemos siempre observar las funciones de autocorrelación para decidir el orden del componente regular (p,q) con base en la evolución de los primeros coeficientes.

En el cuadro 1 se incluyen los procesos AR, MA y ARMA que deberán utilizarse en caso de que la serie estacionaria muestre un comportamiento en la función de autocorrelación y autocorrelación parcial similar a las funciones teóricas de autocorrelación y autocorrelación parcial contenidas en dicho cuadro.

Otro criterio proporcionado por Pulido y López (1999: 291) para determinar los procesos a incluir en el modelo es el siguiente:

Cuadro 11.4.10 Criterio Pulido y López

Función de autocorrelación parcial / Función de autocorrelación	Decrece	Uno o dos coeficientes salen de las bandas de confianza (son significativos)
Decrece	ARMA(1,1)	AR(1) o AR(2)
Uno o dos coeficientes salen de las bandas de confianza (son significativos)	MA(1) o MA(2)	No posible

Finalmente, recuérdese y téngase presente que la forma general de un AR(p) es:

$$Y_t = C + \text{¤}_1 Y_{t-1} + \text{¤}_2 Y_{t-2} + ... + \text{¤}_p Y_{t-p} + a_t$$

Por tanto, un AR(1) tiene la forma:

$$Y_t = C + \text{¤}_1 Y_{t-1} + a_t$$

Mientras que un AR(2):

$$Y_t = C + \text{¤}_1 Y_{t-1} + \text{¤}_2 Y_{t-2} + a_t$$

Por su parte, la forma general de MA(q) es:

$$Y_t = \text{«} + a_t - \text{¯}_1 a_{t-1} - \text{¯}_2 a_{t-2} - ... - \text{¯}_q a_{t-q}$$

Donde a_{t-1} son los residuos.

Así, un proceso MA(1) esta dado por:

$$Y_t = \text{«} + a_t - \text{¯}_1 a_{t-1}$$

Mientras que un MA(2) viene dado por:

$$Y_t = \text{«} + a_t - \text{¯}_1 a_{t-1} - \text{¯}_2 a_{t-2}$$

Y un proceso combinado general ARMA(p,q) está dado por:

$$Y_t = C + \text{¤}_1 Y_{t-1} + \text{¤}_2 Y_{t-2} + ... + \text{¤}_p Y_{t-p} + a_t - \text{¯}_1 a_{t-1} - \text{¯}_2 a_{t-2} - ... - \text{¯}_q a_{t-q}$$

Cabe destacar, como comentan Pulido y López (1999: 281), que “realmente, en las aplicaciones no es frecuente manejar un modelo ARIMA(p,d,q), sino el producto de dos modelos: ARIMA(p,d,q)*SARIMA(P,D,Q), donde el primero corresponde a la parte regular y el segundo a la estacional de la serie.

La existencia de una tendencia en la estacionalidad (a parte de la tendencia regular de la serie) puede venir provocada por una estacionalidad no rígida en que, por ejemplo, las puntas de estacionalidad se agudicen (o reduzcan) con el paso del tiempo.

Ahora bien, para identificar los procesos que se incluirán en el modelo o los modelos posibles a estimar, como se comento, es necesario contar con las funciones de autocorrelación y autocorrelación parcial del correlograma de la serie estacionaria (ver la página 642), Gráfico 11-4.10, en nuestro caso de la serie: DLARA.

Como puede observarse en el correlograma, el rezago séptimo sale de las bandas de confianza tanto de las funciones de autocorrelación como de autocorrelación parcial; asimismo, el rezago 30 sale de las bandas de confianza para cada función; y, el rezago 31 sale de las bandas de confianza en la función de autocorrelación parcial. Ahora comparémos estos valores reales con los modelos teóricos siguientes:

Imagen 11.4.11

Comportamiento teórico para identificar en la función

de autocorrelación total y parcial los procesos AR, MA Y ARMA.

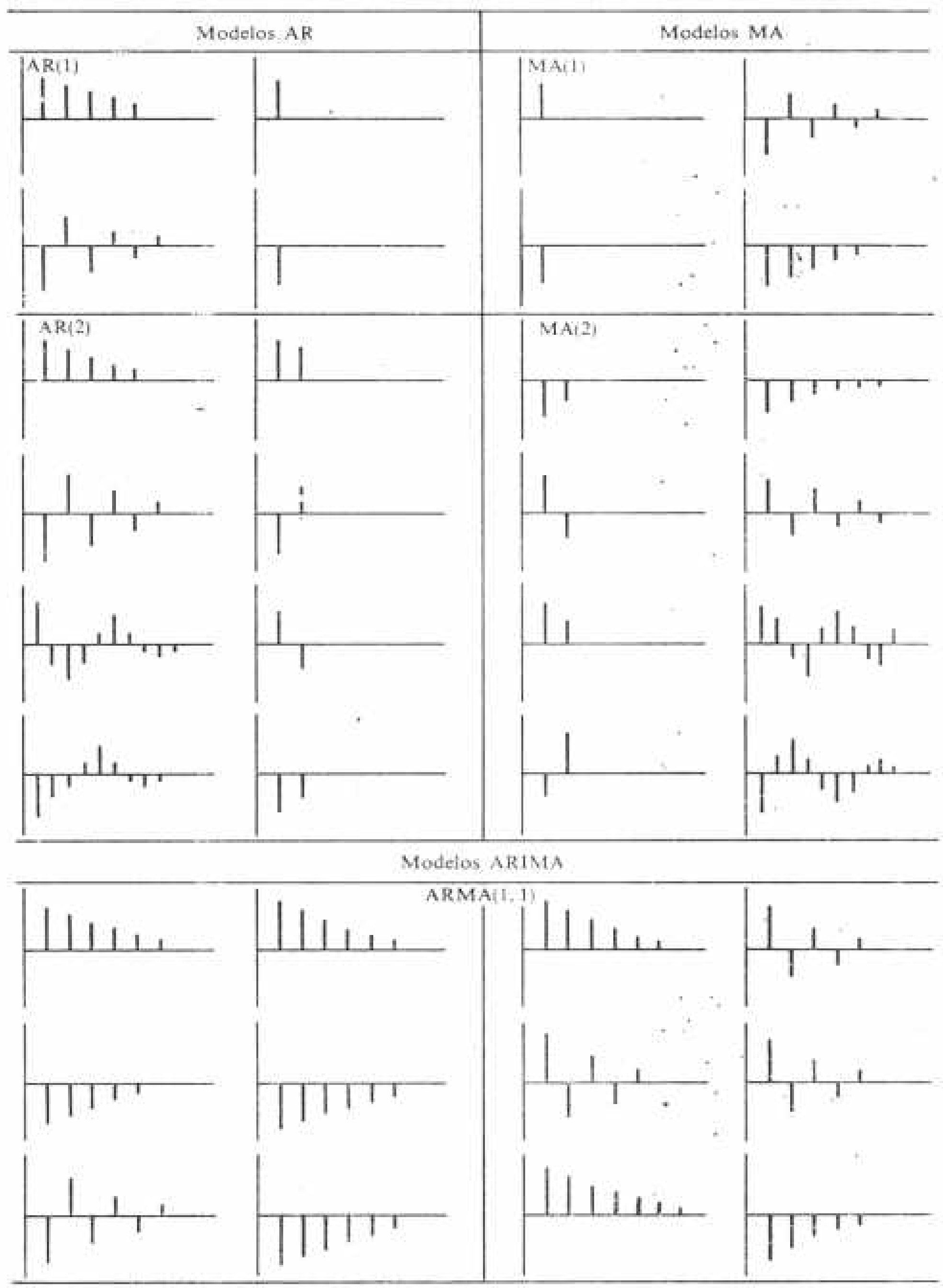


Figura 20.2.—Funciones teóricas de autocorrelación y autocorrelación parcial (izquierda y derecha en cada caso) para diferentes tipos de modelos.

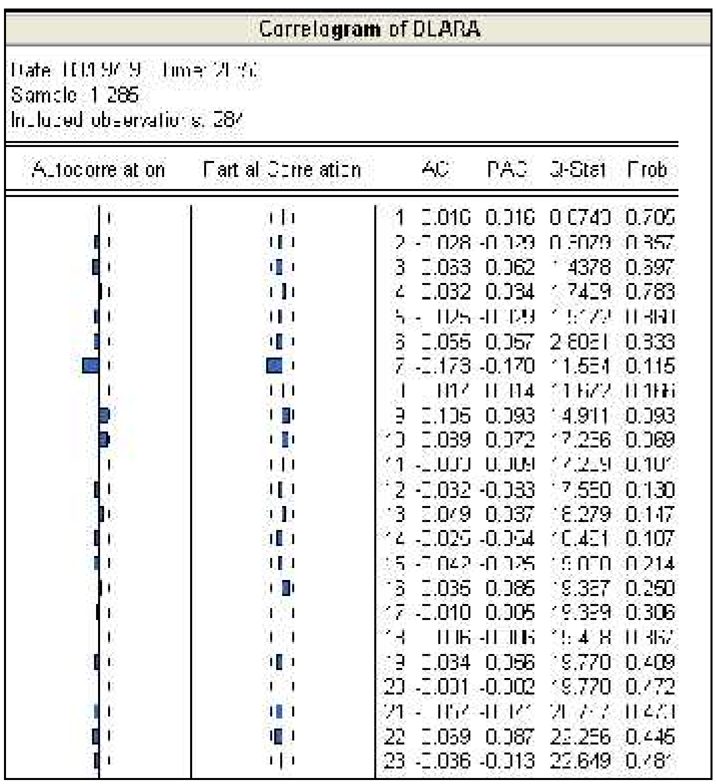
En virtud de que el comportamiento mostrado en el correlograma de la serie DLARA no corresponde con el comportamiento teórico establecido en el cuadro 1, vamos a establecer dos modelos hipotéticos para, posteriormente, de acuerdo a determinados criterios que veremos más adelante, discernir entre uno u otros modelos.

De tal suerte y derivado del análisis visual realizado a través del correlograma para el caso de la serie DLARA y tras realizar diversos modelos con ayuda de Eviews, los modelos identificados que vamos a desarrollar aquí son:

1. AR(7) MA(7), es decir, ARMA(7,7) y en virtud de que diferenciamos una vez la serie para convertirla en estacionaria es integrada de orden 1, I(1). Por tanto, nuestro primer modelo tiene la forma: ARIMA(7,1,7); y

2. AR(7) MA(7) y en la parte estacional un SMA(31), y como diferenciamos una vez para convertir la serie en estacionaria, la serie DLARA es integrada de orden 1, I(1). Por tanto, el segundo modelo a estimar y contrastar es: ARIMA(7,1,7)*SARIMA(0,0,31).

Grafico 11.4.6 Correlograma



Correlogram of DLARA

Date: [illegible] Time: [illegible]
Sample: 1 285
Included observations: 284

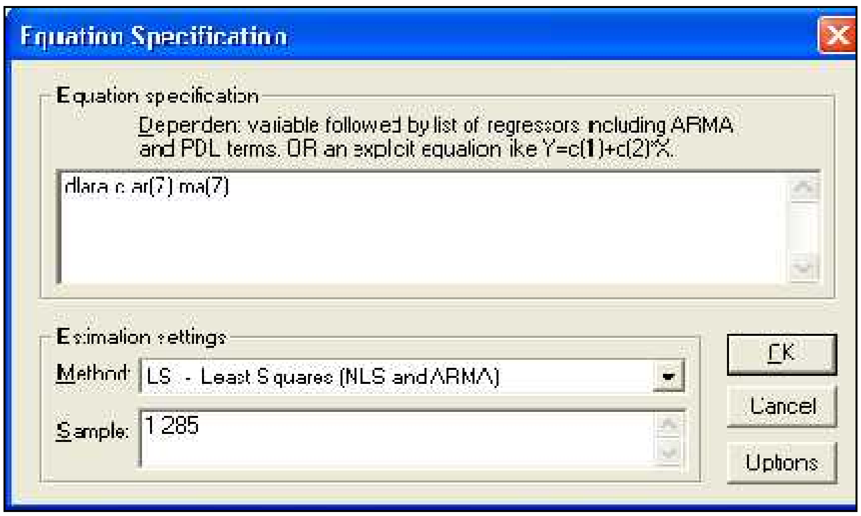
	AC	PAC	Q-Stat	Prob
1	0.016	0.016	0.0740	0.705
2	-0.028	-0.029	0.3079	0.857
3	0.063	0.062	1.4378	0.697
4	0.032	0.034	1.7429	0.783
5	[illegible]	[illegible]	[illegible]	[illegible]
6	0.055	0.057	2.8031	0.833
7	-0.173	-0.170	11.531	0.115
8	[illegible]	[illegible]	[illegible]	[illegible]
9	0.105	0.093	14.911	0.093
10	0.039	0.072	17.236	0.069
11	-0.003	0.009	17.239	0.101
12	-0.032	-0.033	17.550	0.130
13	0.049	0.037	18.279	0.147
14	-0.025	-0.054	18.451	0.107
15	-0.042	-0.025	19.070	0.214
16	0.036	0.085	19.387	0.250
17	-0.010	0.005	19.389	0.306
18	[illegible]	[illegible]	[illegible]	[illegible]
19	0.034	0.056	19.770	0.409
20	-0.001	-0.002	19.770	0.472
21	[illegible]	[illegible]	[illegible]	[illegible]
22	0.069	0.087	22.256	0.446
23	-0.036	-0.013	22.649	[illegible]

IV) Estimación de coeficientes y evaluación-validación del modelo.

La evaluación de modelos ARIMA puede hacerse en forma similar a como se hace con el modelo de regresión o los modelos de ecuaciones simultaneas (Pulido y Pérez, 2000)

Así, una vez identificado el modelo o los modelos procedemos a estimarlos con ayuda de Eviews de la siguiente forma: en el menú principal vamos a: Quick/Estimate Equation/, y en la ventana que despliega escribimos el primer modelo: dlara c ar(7) ma(7)/Ok.

Imagen 11.4.12 Comandos



Con lo cual obtenemos los siguientes resultados:

Cuadro 11.4.11 Regresion

Dependent Variable: DLARA				
Method: Least Squares				
Date: 03/11/09 Time: 13:48				
Sample(adjusted): 9 285				
Included observations: 277 after adjusting endpoints				
Convergence achieved after 24 iterations				
Backcast: 2 8				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004593	0.001342	-3.421214	0.0007
AR(7)	0.653180	0.102730	6.358233	0.0000
MA(7)	-0.820713	0.078910	-10.40057	0.0000
R-squared	0.056671	Mean dependent var		-0.003994
Adjusted R-squared	0.049785	S.D. dependent var		0.038360
S.E. of regression	0.037393	Akaike info criterion		-3.723910
Sum squared resid	0.383111	Schwarz criterion		-3.684661
Log likelihood	518.7615	F-statistic		8.230282
Durbin-Watson stat	2.016104	Prob(F-statistic)		0.000338
Inverted AR Roots	.94	.59+.74i	.59 -.74i	-.21 -.92i
	-.21+.92i	-.85 -.41i	-.85+.41i	
Inverted MA Roots	.97	.61+.76i	.61 -.76i	-.22 -.95i
	-.22+.95i	-.88 -.42i	-.88+.42i	

Con los resultados obtenidos hay que juzgar la validez del modelo estimado, para ello debemos verificar:

- Contraste "t" de significación de coeficientes: Para nuestro caso, al ser la probabilidad asociada a las estadísticas "t" para cada coeficiente menores que 0.05 concluimos que los coeficientes son estadísticamente significativos.

- Matriz de covarianzas: Para acceder a ella dentro del menú de la ecuación vamos a: View/Covariance Matriz/Ok. Obteniendo:

Cuadro 11.4.12 Correlograma

	C	AR(7)	MA(7)
C	1.80212708265e-06	-9.36143365238e-06	5.98178985982e-06
AR(7)	-9.36143365238e-06	0.0105534103408	-0.00724748850113
MA(7)	5.98178985982e-06	-0.00724748850113	0.00622684554418

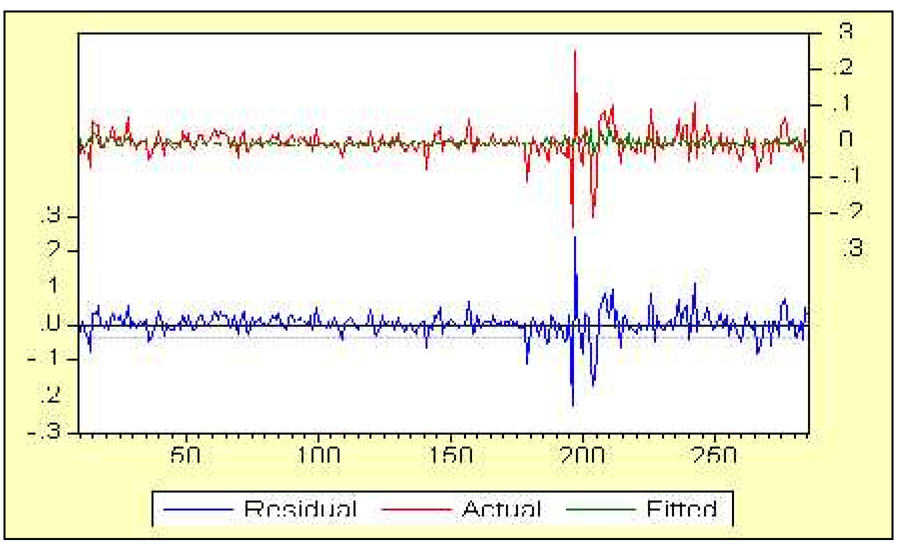
En este caso, como se recordara, lo que nos interesa analizar son los valores fuera de la diagonal principal, ¿Por qué? Así, por un lado, covarianzas relativamente altas indican parámetros en cierta forma redundantes, algunos de los cuales podrían eliminarse sin gran perjuicio para la capacidad explicativa del modelo. Por otro, covarianzas reducidas conforman la conveniencia de mantener todos los coeficientes del modelo (Pulido y López, 1999: 306).

Para nuestro caso, observamos que la covarianza que existe entre los procesos AR(7) y Ma(7) y entre cada uno de estos con la constante es muy reducida, por tanto, concluimos que debemos mantener todos los coeficientes del modelo y podemos seguir adelante.

- <u>Tabla y gráfico de residuos:</u> Para acceder a la tabla que nos muestra los residuos, así como los valores de la serie original y estimada, dentro del menú de la ventana de la ecuación del modelo estimado seleccionamos: View/Actual,fitted, residual table/Ok. Y para acceder el gráfico del ajuste realizado a través del modelo vamos a: View/Actual, fitted, residual graph/Ok.

Para nuestro caso la gráfica de los residuos obtenidos a través del modelo planteado es:

Grafico 11.4.7 Comportamiento de las series

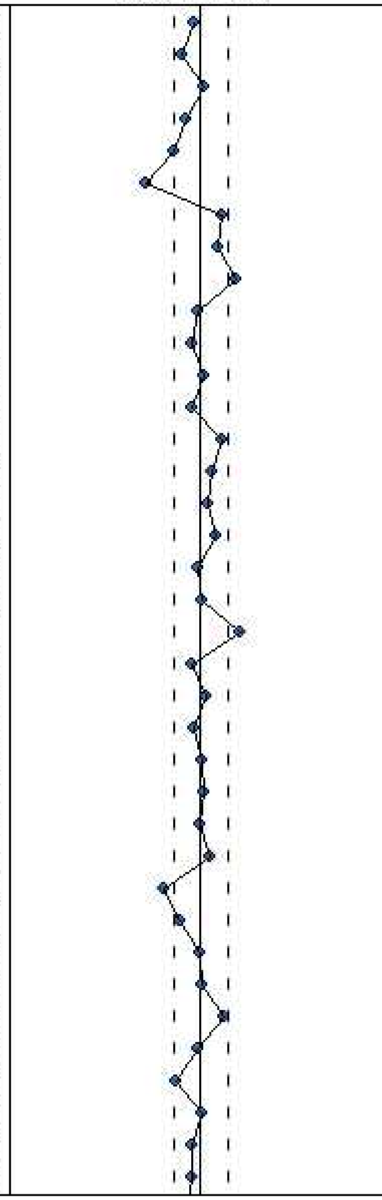


Para el caso de la tabla de los valores reales, los estimados y la gráfica, por razones de espacio tan sólo vamos a mostrar una parte:

Grafco 11.4.8 Valores estimados y residuales

obs	Actual	Fitted	Residual
9	0.02036	0.02872	-0.00836
10	-0.03416	-0.00985	-0.02432
11	-0.01048	-0.01519	0.00471
12	-0.03575	-0.01629	-0.01946
13	-0.03049	0.00483	-0.03531
14	-0.07094	0.00292	-0.07385
15	0.05486	0.02318	0.03168
16	0.04383	0.01857	0.02526
17	0.04635	-0.00395	0.05030
18	-0.01669	-0.01230	-0.00439
19	-0.01968	-0.00898	-0.01070
20	0.01435	0.00748	0.00687
21	0.00000	0.01269	-0.01269
22	0.03672	0.00824	0.02848
23	0.02375	0.00630	0.01745
24	-0.00252	-0.01260	0.01008
25	0.01253	-0.00889	0.02142
26	-0.00833	-0.00566	-0.00267
27	0.00418	0.00214	0.00204
28	0.06532	0.00882	0.05650
29	-0.01257	-0.00098	-0.01159
30	0.00787	-0.00040	0.00827
31	-0.02060	-0.01151	-0.00909
32	-0.00885	-0.01099	0.00214
33	0.00242	-0.00484	0.00726
34	-0.00081	-0.00054	-0.00027
35	0.00723	-0.00530	0.01253
36	-0.05091	-0.00029	-0.05062
37	-0.03254	-0.00324	-0.02930
38	-0.00786	-0.00759	-0.00027
39	-0.00616	-0.00913	0.00297
40	0.02699	-0.00597	0.03297
41	-0.00430	-0.00190	-0.00241
42	-0.04050	-0.00715	-0.03335
43	0.00805	0.00670	0.00136
44	-0.01165	0.00120	-0.01285
45	-0.01912	-0.00650	-0.01261

Residual Plot



Como puede observarse en la gráfica de los valores reales, estimados y residuales, estos últimos se muestran dentro, durante la mayor parte del periodo, de las bandas de confianza, con excepción clara del periodo comprendido entre las observaciones 195 a la 215, periodo de gran volatilidad del precio de las acciones de Grupo ARA. Este periodo comprende a partir del 2 de octubre de

2008 al 6 de noviembre del mismo año, periodo que se enmarca dentro del derrumbe de las bolsas de valores del mundo, es decir, el comportamiento se explica por la incertidumbre y pánico generado a raíz del crack bursátil y por tanto quiebra de diferentes empresas a nivel mundial, entre ellas, por supuesto, las grandes compañías hipotecarias, que sin duda, impactaron desfavorablemente a diferentes empresas del ramo a nivel mundial, entre ellas: Consorcio ARA. De hecho, fue el 23 de octubre de 2008 (observación 204 correspondiente) cuando el precio de las acciones de ARA alcanzaron su precio más bajo durante todo el periodo en cuestión, a saber: 3.52 pesos por acción. Aunque, también es cierto que en las observaciones 196 y 197 el precio real de las acciones de ARA difiere en sobremanera con el precio estimado, de ahí el comportamiento que muestra o presenta la gráfica. Es decir, el error en estas observaciones es alto.

Cabe destacar que el gráfico de los residuos nos proporciona una visión de conjunto de la cuantía de los errores, sesgos sistemáticos y puntos de errores excepcionales (Pulido y López, 1999: 311).

- Prueba F global: Como se sabe la prueba F es un contraste de significación de conjunto de todo el modelo. Así, si la probabilidad asociada al estadístico F es menor que 0.05 se concluye que las variables independientes en su conjunto explican el comportamiento de la variable dependiente y viceversa. En nuestro caso, al ser la probabilidad asociada a F menor que 0.05 concluimos que el modelo en su conjunto es adecuado.

- Coeficiente de determinación (R^2): En el análisis de series de tiempo es suficiente que el R^2 supere ampliamente los valores de 0.2 o 0.4 siempre y cuando se trabaje con un proceso estocástico en diferencias. De acuerdo a Pulido y López (1999: 311) “un coeficiente de determinación de 0.5 a 0.7 en una variable en diferencias puede ser frecuentemente equivalente o superior a 0.9 en la variable original”.

En nuestro caso, tenemos que $R^2 = 0.056671$, lo cual, no cumple con lo especificado por Pulido y López; sin embargo, téngase presente que ellos hablan tan sólo de variables en diferencias, mientras que nosotros estamos trabajando la serie en logaritmos diferenciada.

- Correlograma de los residuos: Siguiendo con la validación del modelo, vamos a observar el correlograma de los residuos obtenidos a partir del modelo planteado. Así, estando en la ventana de la ecuación vamos a: View/Residual Tests/Correlogram-Q-statistics/, mismo que por default nos da 34 rezagos/Ok., obteniendo el siguiente correlograma (ver página 658: Gráfico11.4.9).

Como puede observarse en el correlograma de los residuos, en el rezago 7, tanto en la función de autocorrelación (AC) como en la función de autocorrelación parcial (PAC), ya no salen de las bandas de confianza. Asimismo, en la AC el rezago 30 ya no sale de las bandas de confianza, de hecho se encuentra en el límite. Sin embargo, en la PAC tanto el rezago 30 como 31 salen, aunque poco, de las bandas de confianza. Por otra parte, el valor de los coeficientes de AC y PAC son muy menores a 1. Pero como se comento, los valores para los rezagos 30 y 31 son los más altos de los 34 rezagos mostrados en el correlograma. A pesar de ello, vemos que la probabilidad asociada al estadístico Q son todas mayores a 0.05.

Derivado del análisis realizado al correlograma podemos concluir por adelantado que los residuos son ruido blanco, es decir, $\hat{\rho}_k$ ~ N(0, 1/n).

- <u>Test DF, ADF y PP para los residuos:</u> Para verificar que los residuos son ruido blanco, podemos ahora realizar sobre ellos los Test DF, ADF Y PP. Para ello, debemos crear la serie de los residuos. Así, volvemos a correr la regresión, para ello tenemos dos opciones: 1) Si estamos en la ventana de la ecuación vamos a: Estimate/Ok.; y 2) En el menú principal vamos a: Quick/Estimate Equation…/y escribimos la ecuación planteada con antelación. El porque hay que realizar estos pasos se debe a que el programa tan sólo guarda la información en tanto que no se realice una nueva operación, como es el caso de una nueva regresión, es decir, guarda la información momentáneamente. De tal suerte que, al correr la regresión o al volverla a estimar Eviews generará los residuos de la misma automáticamente en la serie “resid” que aparece en el Workfile. A continuación en el menú del Workfile vamos a: Objects/New Object…/Series/, y le damos el nombre de U1/Ok. Misma que aparecerá en la ventana del archivo de trabajo. De la serie “Resid” generada momentáneamente copiamos los valores de los residuos a partir de la observación 2 hasta la 285 y las copiamos en la nueva serie creada denominada U1 habiendo oprimido con antelación el botón “Edit +/-“ de la ventana de la nueva serie para que nos permita incluir los datos. Una vez hecho esto, podemos ahora realizar las pruebas de raíces unitarias sobre los residuos para probar si son o no ruido blanco.

Grafico 11.4.9 Correlograma

Correlogram of Residuals

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.009	-0.009	0.0214	
		2	-0.032	-0.033	0.3168	
		3	-0.028	-0.029	0.5431	0.461
		4	0.045	0.043	1.1100	0.574
		5	-0.033	-0.034	1.4148	0.702
		6	-0.068	-0.067	2.7407	0.602
		7	-0.045	-0.046	3.3245	0.650
		8	-0.012	-0.021	3.3629	0.762
		9	0.081	0.077	5.2375	0.631
		10	0.100	0.104	8.1142	0.422
		11	0.002	0.009	8.1158	0.523
		12	-0.034	-0.031	8.4514	0.585
		13	0.026	0.016	8.6430	0.655
		14	0.048	0.040	9.3095	0.676
		15	-0.068	-0.052	10.676	0.638
		16	0.024	0.051	10.846	0.698
		17	0.021	0.030	10.981	0.754
		18	0.002	-0.008	10.982	0.811
		19	-0.048	-0.054	11.658	0.820
		20	-0.007	-0.021	11.674	0.864
		21	-0.009	-0.012	11.698	0.898
		22	-0.088	-0.088	14.040	0.828
		23	-0.035	-0.042	14.414	0.851
		24	0.014	0.008	14.474	0.884
		25	-0.004	-0.010	14.479	0.912
		26	0.032	0.024	14.795	0.927
		27	0.004	-0.011	14.800	0.946
		28	0.070	0.069	16.336	0.928
		29	-0.017	-0.008	16.424	0.944
		30	-0.123	-0.137	21.182	0.818
		31	0.114	0.140	25.277	0.664
		32	-0.088	-0.072	27.736	0.584
		33	-0.033	-0.020	28.074	0.617
		34	0.004	0.019	28.078	0.665

Como ya se sabe, para realizar estas pruebas estando en la ventana de la serie, en este caso U1, vamos a: View/Unit root test/, y seleccionamos tanto los test DF, ADF y PP. Tan sólo indicaremos que en virtud de que la serie de los residuos para esta regresión no presenta tendencia indicaremos para los tres test “None”, obteniendo los siguientes resultados:

- Para DF:

Cuadro 11.4.13 Prueba Dickey Fuller

Augmented Dickey-Fuller Unit Root Test on U1

Null Hypothesis: U1 has a unit root
Exogenous: None
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-16.91242	0.0000
Test critical values:	1% level	-2.573159	
	5% level	-1.941949	
	10% level	-1.615950	

*MacKinnon (1996) one-sided p-values.

- Para ADF:

Cuadro 11.4.14 Prueba Dickey Fuller Aumentada

Augmented Dickey-Fuller Unit Root Test on U1

Null Hypothesis: U1 has a unit root
Exogenous: None
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-12.27466	0.0000
Test critical values:	1% level	-2.573188	
	5% level	-1.941953	
	10% level	-1.615948	

*MacKinnon (1996) one-sided p-values.

- Para PP:

Cuadro 11.4.15 Prueba Phillips Perron

Phillips-Perron Unit Root Test on U1

Null Hypothesis: U1 has a unit root
Exogenous: None
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-16.91238	0.0000
Test critical values:	1% level	-2.573159	
	5% level	-1.941949	
	10% level	-1.615950	

*MacKinnon (1996) one-sided p-values.

Conclusión: Como en las tres pruebas la probabilidad asociada a cada estadístico es menor a 0.05 y en virtud de que el valor de los estadísticos DF, ADF y PP son mayores a los valores críticos de MacKinnon al 5% concluimos que la serie de los residuos no presenta raíz unitaria, por tanto, es ruido blanco, es decir, tiene un comportamiento aleatorio.

Ahora realicemos el segundo modelo planteado siguiendo y/o aplicando los mismos pasos y criterios descritos en el punto IV para el primer modelo. Por razones de espacio, tan sólo mostraremos los resultados obtenidos dejando al lector la interpretación, misma que puede realizarse como práctica. De tal suerte que los resultados son los siguientes:

- Resultados de la ecuación de regresión:

Cuadro 11.4.16 Regresion

Dependent Variable: DLARA
Method: Least Squares
Date: 03/11/09 Time: 15:56
Sample(adjusted): 9 285
Included observations: 277 after adjusting endpoints
Convergence achieved after 18 iterations
Backcast: -29 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.005370	0.001333	-4.029194	0.0001
AR(7)	0.810983	0.039580	20.48957	0.0000
MA(7)	-0.949112	0.016044	-59.15614	0.0000
SMA(31)	0.127127	0.063141	2.013397	0.0451

R-squared	0.073251	Mean dependent var	-0.003994
Adjusted R-squared	0.063067	S.D. dependent var	0.038360
S.E. of regression	0.037130	Akaike info criterion	-3.734423
Sum squared resid	0.376377	Schwarz criterion	-3.682090
Log likelihood	521.2175	F-statistic	7.192729
Durbin-Watson stat	1.969314	Prob(F-statistic)	0.000116

Inverted AR Roots	.97	.61 -.76i	.61+.76i	-.22 -.95i
	-.22+.95i	-.87+.42i	-.87 -.42i	
Inverted MA Roots	.99	.93 -.09i	.93+.09i	.89+.28i
	.89 -.28i	.82 -.45i	.82+.45i	.71+.61i
	.71 -.61i	.62+.78i	.62 -.78i	.57 -.74i
	.57+.74i	.41+.84i	.41 -.84i	.23 -.91i
	.23+.91i	.05 -.93i	.05+.93i	-.14+.92i
	-.14 -.92i	-.22+.97i	-.22 -.97i	-.32+.88i
	-.32 -.88i	-.49+.79i	-.49 -.79i	-.64+.68i
	-.64 -.68i	-.77+.53i	-.77 -.53i	-.86+.37i
	-.86 -.37i	-.89 -.43i	-.89+.43i	-.92+.19i
	-.92 -.19i	-.94		

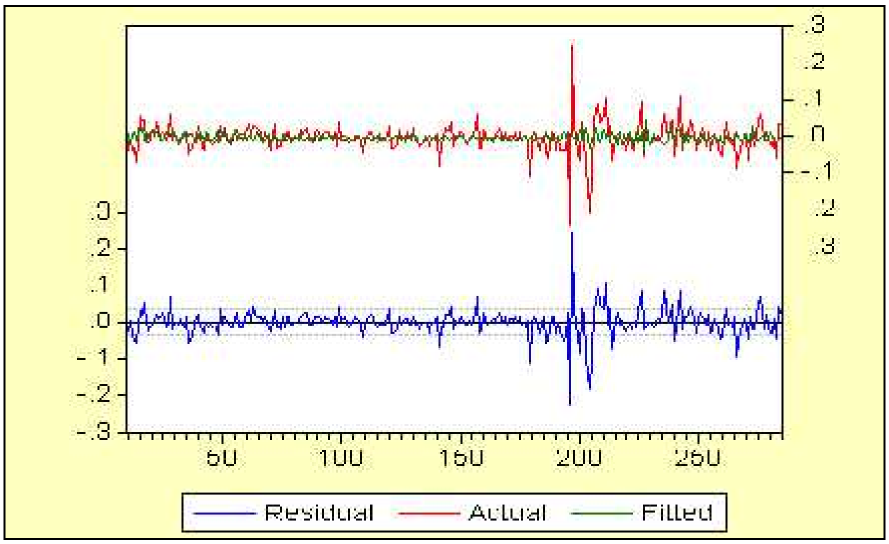
- Matriz de covarianzas:

Cuadro 11.4.17 Correlograma

	C	AR(7)	MA(7)	SMA(31)
C	1.77660057177e-06	-6.74619695222e-06	9.79654783032e-07	-2.51274973952e-06
AR(7)	-6.74619695222e-06	0.00156660030031	-0.000299034451301	-0.000160631335892
MA(7)	9.79654783032e-07	-0.000299034451301	0.000257415768545	1.67250225397e-05
SMA(31)	-2.51274973952e-06	-0.000160631335892	1.67250225397e-05	0.00398672737268

- <u>Tabla y gráfica de los residuos:</u>

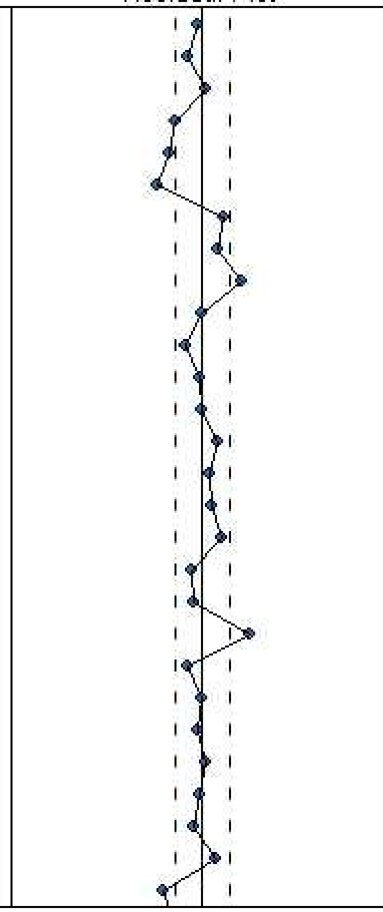
Grafico 11.4.10 Comportamiento de las series



Cuadro 11.4.18 Valores estimados y errores

obs	Actual	Fitted	Residual
9	0.02036	0.02642	-0.00607
10	-0.03416	-0.01294	-0.02122
11	-0.01048	-0.01724	0.00676
12	-0.03575	-3.8E-05	-0.03571
13	-0.03049	0.01547	-0.04595
14	-0.07094	-0.01076	-0.06018
15	0.05486	0.02321	0.03166
16	0.04383	0.02177	0.02206
17	0.04635	-0.00735	0.05370
18	-0.01669	-0.01616	-0.00053
19	-0.01968	0.00407	-0.02375
20	0.01435	0.01845	-0.00410
21	0.00000	-0.00049	0.00049
22	0.03672	0.01333	0.02339
23	0.02375	0.01343	0.01032
24	-0.00252	-0.01509	0.01257
25	0.01253	-0.01361	0.02614
26	-0.00833	0.00571	-0.01404
27	0.00418	0.01447	-0.01029
28	0.06532	-0.00044	0.06576
29	-0.01257	0.00583	-0.01840
30	0.00787	0.00819	-0.00032
31	-0.02060	-0.01526	-0.00534
32	-0.00885	-0.01598	0.00713
33	0.00242	0.00397	-0.00155
34	-0.00081	0.01116	-0.01196
35	0.00723	-0.01098	0.01821
36	-0.05091	0.00291	-0.05382

Residual Plot



- Correlograma de los residuos:

Cuadro 11.4.19 Correlograma

Correlogram of Residuals

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.014	0.014	0.0585	
		2	-0.037	-0.037	0.4368	
		3	-0.045	-0.044	1.0010	
		4	0.036	0.036	1.3598	0.244
		5	-0.044	-0.049	1.9126	0.384
		6	-0.083	-0.082	3.8838	0.274
		7	-0.087	-0.086	6.0583	0.195
		8	-0.006	-0.016	6.0675	0.300
		9	0.085	0.076	8.1502	0.227
		10	0.094	0.090	10.711	0.152
		11	0.011	0.014	10.748	0.216
		12	-0.020	-0.022	10.860	0.285
		13	0.005	-0.008	10.867	0.368
		14	0.020	0.012	10.988	0.444
		15	-0.069	-0.053	12.375	0.416
		16	0.033	0.068	12.704	0.471
		17	0.031	0.047	12.987	0.528
		18	0.006	-0.002	12.998	0.602
		19	-0.039	-0.048	13.463	0.639
		20	-0.028	-0.046	13.704	0.688
		21	-0.024	-0.034	13.883	0.737
		22	-0.103	-0.108	17.111	0.582
		23	-0.041	-0.032	17.627	0.612
		24	0.020	0.027	17.751	0.665
		25	0.019	0.007	17.861	0.714
		26	0.046	0.028	18.522	0.729
		27	-0.006	-0.035	18.533	0.776
		28	0.086	0.072	20.855	0.701
		29	-0.002	-0.007	20.857	0.749
		30	-0.136	-0.141	26.611	0.485
		31	-0.009	0.035	26.637	0.538
		32	-0.091	-0.077	29.272	0.451
		33	-0.033	-0.029	29.625	0.485
		34	-0.010	-0.009	29.659	0.535
		35	[illegible]	[illegible]	[illegible]	[illegible]

<u>Comentario sobre el correlograma:</u> Como puede observarse en las funciones de autocorrelación (AC) y autocorrelación parcial (PAC) ninguna sale de las bandas de confianza, pues en el rezago 30 ambas están en el límite; asimismo, los valores de AC y PAC son muy menores a 1; y la probabilidad asociada al estadístico Q son mayores que 0.05. Derivado de estos resultados podemos concluir que los residuos son ruido blanco.

- Test DF, ADF y PP para los residuos (en ese orden):

Cuadro 11.4.20 Prueba Dickey Fuller

Augmented Dickey-Fuller Unit Root Test on U2

Null Hypothesis: U2 has a unit root
Exogenous: None
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-16.47504	0.0000
Test critical values:	1% level	-2.573188	
	5% level	-1.941953	
	10% level	1.615948	

*MacKinnon (1996) one-sided p-values.

Cuadro 11.4.21 Prueba Dickey Fuller Aumentada

Augmented Dickey-Fuller Unit Root Test on U2

Null Hypothesis: U2 has a unit root
Exogenous: None
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-12.16534	0.0000
Test critical values:	1% level	-2.573217	
	5% level	-1.941957	
	10% level	-1.615945	

*MacKinnon (1996) one-sided p-values.

Cuadro 11.4.22 Prueba Phillips Perron

Phillips-Perron Unit Root Test on U2			
Null Hypothesis: U2 has a unit root			
Exogenous: None			
Bandwidth: 1 (Fixed using Bartlett kernel)			
		Adj. t Stat	Prob.*
Phillips-Perron test statistic		-16.47505	0.0000
Test critical values:	1% level	-2.573188	
	5% level	-1.941953	
	10% level	-1.615948	
*MacKinnon (1996) one-sided p-values.			

V) Selección del modelo.

Ahora bien, si varios modelos pasan las pruebas de validación y se quiere discernir entre uno u otros, existen criterios para determinar cual de ellos es mejor. Es decir, los siguientes criterios se utilizan principalmente como elemento de comparación entre modelos alternativos para una misma variable, entre ellos tenemos:

1. Función logarítmica de verosimilitud (log likelihood):

Viene dado por:

$$L = \frac{-I}{2}\left[1 + \ln(2\text{®}) + \ln\frac{\Sigma e_i^2}{I}\right]$$

Donde:

I = Número de observaciones.

Σe_i^2 = Sumatoria de los errores de la regresión al cuadrado.

De acuerdo a Cabrera, *et al.*- (2001: 47) el valor que toma la función logarítmica de verosimilitud puede ser un criterio para valorar la bondad del modelo, esencialmente cuando se trata de comparar dos o más modelos alternativos. Pero, ¿Por qué no utilizar el R^2 para discernir entre un modelo u otro? Porque la principal ventaja de este estadístico frente al R^2 es que no se ve influido por las transformaciones a que se puede ver sometida la variable endógena. Por lo que se refiere a este criterio para la elección entre distintos modelos, se escoge aquel modelo que presenta un valor de la función logarítmica de verosimilitud (L) más alto.

2. <u>Criterio de información de Akaike (AIC: Akaike info criterion):</u>

Esta dado por:

$$AIC = \frac{2K}{I} - \frac{2L}{I}$$

Donde:

K= Número de variables exógenas incluidas en el modelo.

L = Valor de la función logarítmica de verosimilitud.

I = Número de observaciones.

El AIC sirve para comparar la bondad de ajuste entre dos modelos. El criterio para la elección entre distintos modelos se fundamenta en el valor estimado del estadístico. Se selecciona aquel modelo que presenta un estadístico AIC menor.

3. <u>Criterio de información Schwarz (Schwarz criterion):</u>

Esta dado por:

$$SC = \frac{K * \ln(I)}{I} - \frac{2L}{I}$$

Donde:

K= Número de variables exógenas incluidas en el modelo.

L = Valor de la función logarítmica de verosimilitud.

I = Número de observaciones.

Al igual que L y AIC, éste sirve para comparar la bondad de ajuste entre modelos alternativos. El criterio de elección de un modelo entre varios alternativos se fundamenta en el valor estimado del estadístico. Se selecciona aquel modelo que presenta un estadístico Schwarz menor.

También, para elegir entre un modelo y otro pueden considerarse la suma de cuadrados de los residuos (Sum squared resid) y el error estándar de la regresión (S. E. of regresión). Nos quedaremos con aquel modelo que presente la suma de los errores al cuadrado y el error estándar de la regresión menores. Así mismo, a pesar del problema de utilizar el R^2 como una prueba de la bondad de ajuste del modelo, dado que éste se ve fuertemente influenciado por las transformaciones de la serie, cuando se trata de discernir entre varios modelos nos quedaremos con aquel que posea el R^2 más alto.

De tal suerte, los resultados obtenidos en los dos modelos planteados son:

Tabla 11.4.2 Criterios

Criterios	Primer modelo: ARIMA(7,1,7)	Segundo modelo: ARIMA(7,1,7)*SARIMA(0,0,31)
Suma de cuadrados de los residuos	0,383111	0,376377
Desviación estándar de la regresión	0,037393	0,03713
Coeficiente de determinación	0,056671	0,073251
Coeficiente de determinación ajustado	0,049785	0,063067
Logaritmo de verosimilitud (L)	518,7615	521,2175
Criterio de información de Akaike (AIC)	-3,72391	-3,734423
Criterio de Schwarz (SC)	-3,684661	-3,68209

En virtud de estos resultados y de acuerdo a lo establecido previamente para la selección de un modelo, no cabe duda de que el mejor modelo para estimar el precio de las acciones de Consorcio ARA es el segundo, a saber: ARIMA(7,1,7)*SARIMA(0,0,31). Pues, como puede observarse en el cuadro anterior: la suma de cuadrados de los residuos en el segundo modelo es menor que en el primero; la desviación estándar de la regresión también es menor en el segundo; el segundo modelo presenta un R^2 y $\bar{R}^2$ mayor que el primero; asimismo el segundo modelo presenta un mayor valor de L; el segundo presenta un menor valor de AIC; y en el caso del SC el primer modelo presenta un valor menor que el segundo. Dado que el segundo modelo cumple con la mayor parte de los criterios establecidos, no cabe duda sobre que modelo utilizar.

Antes de seguir adelante es preciso establecer la estructura del modelo. En nuestro caso, el modelo tiene la siguiente estructura:

DLARA = -0.005370 + 0.810983*DLARA(-7) + $Residuo_t$ – (-0.949112)*Residuo(-7) – 0.127127*Residuo(-31) - (-0.949112)*(0.127127*Residuo(-38))

$DLARA_t$ = -0.005370 + 0.810983*DLARA(-7) + $Residuo_t$ + 0.949112*Residuo(-7) – 0.127127*Residuo(-31) + 0.949112*(0.127127*Residuo(-38))

La interpretación de la ecuación anterior es que el precio de las acciones de Consorcio ARA se explica en función de si misma 7 días antes (hábiles o de operación del mercado de valores mexicano), así como por el error que comete el modelo 7 y 31 días anteriores.

Y con esta ecuación de regresión es con la que vamos a proceder a predecir el precio de las acciones de ARA.

VI) Predicción.

Una vez seleccionado y planteado el modelo definitivo, es decir, aquel que cumple satisfactoriamente los criterios de evaluación establecidos, puede pasarse a la etapa de predicción.

La predicción puede realizarse en dos formas diferentes:

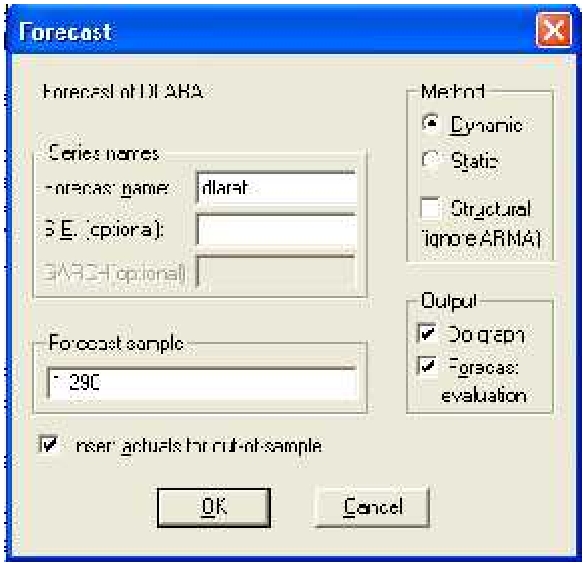
1.- De forma estática o paso a paso; y

2.- De forma dinámica o en cadena.

La diferencia entre una y otra es que la primera nos permite predecir sólo un periodo por delante y utiliza los valores reales para estimar y predecir, por ejemplo, para nuestro caso, tan sólo nos proporcionará el precio de las acciones para la observación 286; mientras que la segunda permite al modelo que vaya realimentando sus propias predicciones, es decir, permite predecir para varias observaciones de una sola vez. Ambas predicciones (estática y dinámica) coinciden para el primer periodo, pero, a partir del segundo, la predicción dinámica utiliza el valor estimado y no el valor real del periodo precedente.

Para nuestro caso, tal y como hemos venido trabajando o más bien dicho, generando la serie, Eviews no la reconoce como transformación y diferenciación de la serie ordinaria (original) ARA, es decir, por ejemplo, a la serie DLARA la reconoce como una serie ordinaria (original) y no como una transformación de la serie ARA. Por lo que, al realizar la predicción, Eviews tan sólo nos permitirá realizarla para la serie DLARA y no para la serie ARA. Por ejemplo, al realizar la predicción de la serie DLARA tal y como hasta el momento hemos venido trabajando, Eviews nos desplegará el siguiente cuadro de diálogo:

Imagen 11.4.13 Prediccion



Como puede observarse, Eviews nos desplegará una ventana para la predicción de DLARA teniendo nosotros que realizar las transformaciones necesarias para obtener los valores de la predicción de la serie ARA.

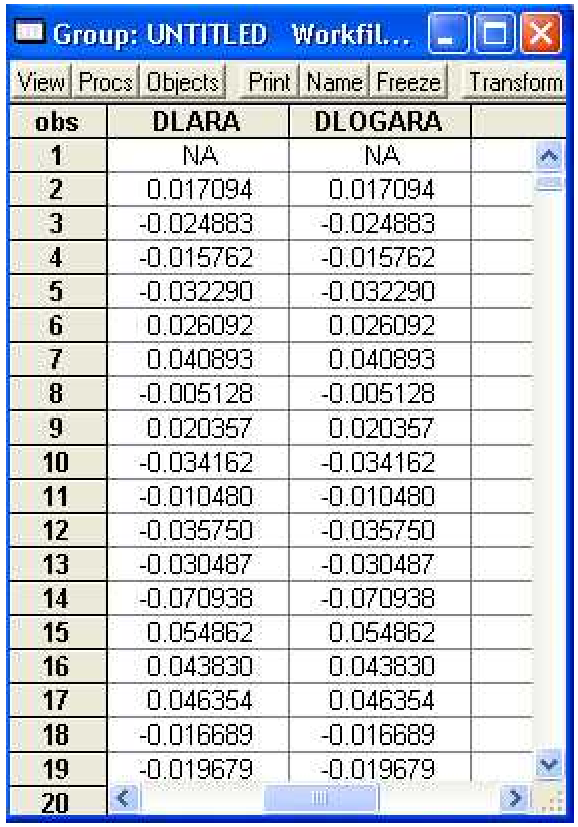
Por tanto, en virtud de que Eviews no reconoce la serie DLARA como una transformación y diferenciación de la serie ordinaria (original) ARA hay que generar una nueva serie. Así, en el menú principal vamos a: Quick/Generate Series…/Ok, y en la venta que despliega escribimos:

dlogara=dlog(ara)

en donde, nuevamente, estamos generando la variable DLARA, pero en términos de que Eviews la reconozca como una serie transformada de la serie original ARA. Entonces, la indicación “dlog(ara)” está generando la primera diferencia de la serie en logaritmos del precio de las acciones de ARA. En este contexto, Eviews reconocerá a la serie ARA como una serie ordinaria y la serie DLOG(ARA) la reconocerá como una transformación y diferenciación de la serie ordinaria ARA. Con ello, cuando realicemos la predicción Eviews nos proporcionará dos opciones de predicción, a saber: 1) En términos de la serie ordinaria, es decir, ARA y por tanto en pesos por acción; y 2) En términos de la serie transformada y diferenciada, es decir, DLOG(ARA) teniendo nosotros que realizar los cálculos necesarios que nos permitan obtener el precio de las acciones de ARA. Nótese que siempre, de aquí en adelante, vamos a trabajar con la serie DLOG(ARA) y no con DLOGARA. Sin embargo, ésta última es la que aparecerá en el Workfile. Pero no nos preocupemos, ya que Eviews reconoce ésta serie aunque no aparezca en el Workfile. Además, siempre podemos trabajar en la barra de comandos.

Así, una vez generada la nueva serie [DLOG(ARA)], misma que por cierto deberá ser exactamente igual a la serie DLARA como puede observarse en el siguiente cuadro:

Imagen 11.4.14 Datos



obs	DLARA	DLOGARA
1	NA	NA
2	0.017094	0.017094
3	-0.024883	-0.024883
4	-0.015762	-0.015762
5	-0.032290	-0.032290
6	0.026092	0.026092
7	0.040893	0.040893
8	-0.005128	-0.005128
9	0.020357	0.020357
10	-0.034162	-0.034162
11	-0.010480	-0.010480
12	-0.035750	-0.035750
13	-0.030487	-0.030487
14	-0.070938	-0.070938
15	0.054862	0.054862
16	0.043830	0.043830
17	0.046354	0.046354
18	-0.016689	-0.016689
19	-0.019679	-0.019679
20		

En el menú principal vamos a: Quick/Estimate Equation…/Ok., y escribimos:

dlog(ara)_c_ar(7)_ma(7)_sma(31)/Ok.

Donde “_” significa espacio.

Obteniendo el siguiente cuadro de resultados:

Cuadro 11.4.23 Regresion

Dependent Variable: DLOG(ARA)
Method: Least Squares
Date: 03/16/09 Time: 18:58
Sample(adjusted): 9 285
Included observations: 277 after adjusting endpoints
Convergence achieved after 18 iterations
Backcast: -29 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.005370	0.001333	-4.029194	0.0001
AR(7)	0.810983	0.039580	20.48957	0.0000
MA(7)	-0.949112	0.016044	-59.15614	0.0000
SMA(31)	0.127127	0.063141	2.013397	0.0451

R-squared	0.073251	Mean dependent var	-0.003994
Adjusted R-squared	0.063067	S.D. dependent var	0.038360
S.E. of regression	0.037130	Akaike info criterion	-3.734423
Sum squared resid	0.376377	Schwarz criterion	-3.682090
Log likelihood	521.2175	F-statistic	7.192729
Durbin-Watson stat	1.969314	Prob(F-statistic)	0.000116

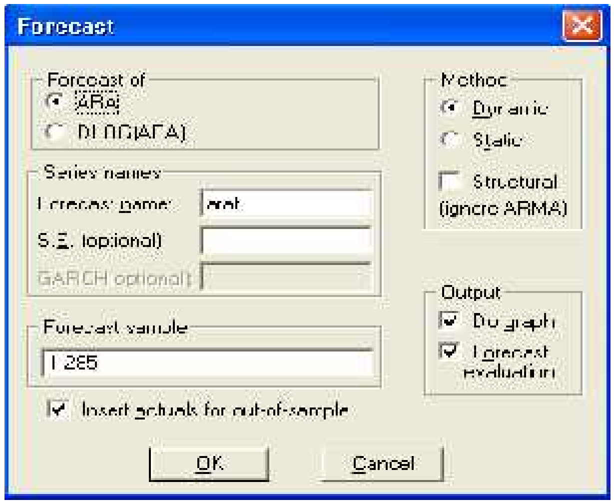
Inverted AR Roots	.97	.61 -.76i	.61+.76i	-.22 -.95i
	-.22+.95i	-.87+.42i	-.87 -.42i	
Inverted MA Roots	.99	.93 -.09i	.93+.09i	.89+.28i
	.89 -.28i	.82 -.45i	.82+.45i	.71+.61i
	.71 -.61i	.62+.78i	.62 -.78i	.57 -.74i
	.57+.74i	.41+.84i	.41 -.84i	.23 -.91i
	.23+.91i	.05 -.93i	.05+.93i	-.14+.92i
	-.14 -.92i	-.22+.97i	-.22 -.97i	-.32+.88i
	-.32 -.88i	-.49+.79i	-.49 -.79i	-.64+.68i
	-.64 -.68i	-.77+.53i	-.77 -.53i	-.86+.37i
	-.86 -.37i	-.89 -.43i	-.89+.43i	-.92+.19i
	-.92 -.19i	-.94		

El mismo resultado puede obtenerse desde la barra de comandos, en la cual escribiremos:

ls_dlog(ara)_c_ar(7)_ma(7)_sma(31)/Ok.

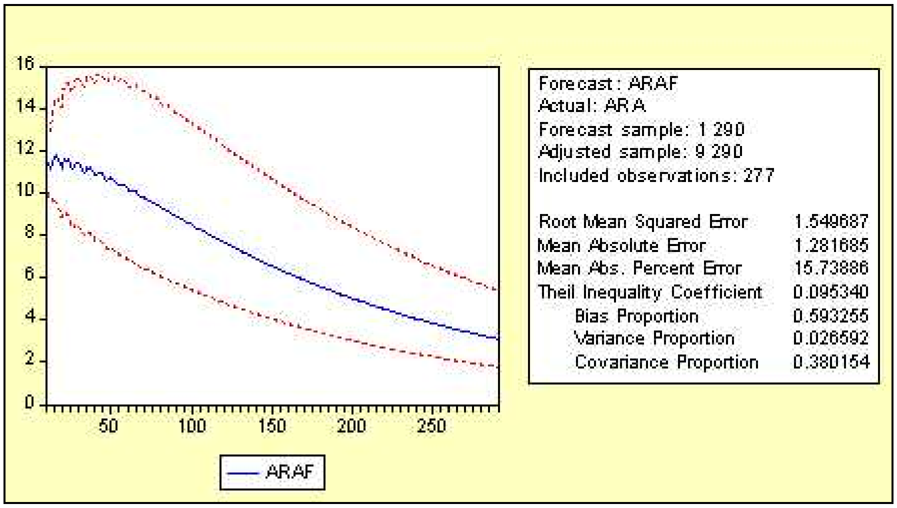
En donde: “ls” indica estimación por mínimos cuadrados y “_” espacio.

Imagen 11.4.15 Comando



En la ventana de la ecuación presionamos el botón “FORECAST”, lo cual desplegará una ventana como la anterior. Como puede observarse en esta ventana a diferencia de la primera (ver página 38), Eviews nos da a escoger entre dos términos de predicción (Forecast of), a saber: 1) En términos de la serie ARA; y 2) En términos de la serie transformada y diferenciada de ARA [DLOG(ARA)], en nuestro caso seleccionaremos ARA. Luego, nos solicita el nombre de la serie que predecirá (Series names), el cual por default Eviews nos da: araf, donde la “f” indica la predicción, es decir, forecast. En seguida, el programa nos solicita el tamaño de la muestra para la predicción (Forecast sample); el cual para nuestros fines modificaremos de 1 a 290. Y, también nos solicita el método de predicción (Method), es decir, si la predicción será dinámica (Dynamic) o estática (Static); para nuestro caso, en virtud de que nos interesa predecir para la observación 286 a 290, seleccionaremos el método dinámico. El resto lo dejamos tal cual lo específica el programa por default/Ok. La nueva variable aparecerá en el Workfile y asimismo, nos desplegará una gráfica que muestra estadísticos que permiten determinar la bondad de la predicción. Así, la gráfica es

Grafico11.4.10 Prediccion



Como puede observarse, la serie ARAF está dentro de las bandas de confianza por lo que podemos concluir que la predicción es buena. Asimismo, con respecto a las estadísticas, dado que casi todas (con la excepción de root mean squared error, mean absolute error y mean abs, percent error) tienen valores menores a uno, decimos que es buena estimación de la predicción.

El siguiente cuadro muestra los precios reales de las acciones de ARA para la observación 281 a 290, la estimación y predicción realizada con nuestro modelo con el método dinámico para las mismas observaciones, así como el error de predicción:

Tabla 11.4.3 Datos

Obs.	Fecha	Precio real (Pr)	Precio estimado (Pe)	Error (Pr-Pe)
281	13/Feb/2009	3.87	3.25	0.62
282	16/Feb/2009	3.86	3.23	0.63
283	17/Feb/2009	3.64	3.21	0.43
284	18/Feb/2009	3.76	3.20	0.56
285	19/Feb/2009	3.86	3.18	0.68
286	20/Feb/2009	3.71	3.16	0.55
287	23/Feb/2009	3.67	3.15	0.52
288	24/Feb/2009	3.75	3.13	0.62
289	25/Feb/2009	3.65	3.11	0.54

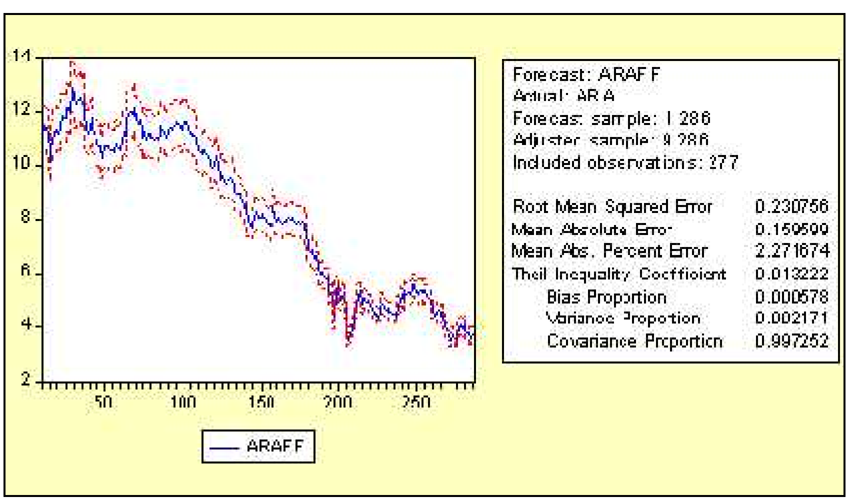
290	26/Feb/2009	3.83	3.09	0.74

Como puede observarse, el error de estimación (para el caso de las observaciones 281 a 285) y el error de predicción (para el caso de las observaciones 286 a 290) es grande, en promedio de 0.589 pesos por acción, con el método dinámico. Así, la predicción nos proporciona una guía para determinar, en este caso, la compra o venta de acciones de Consorcio ARA. Téngase presente, además, que los modelos ideales para predecir series financieras son los ARCH y GARCH. Sin embargo, puede utilizarse la metodología Box & Jenkins para este tipo de series recomendándose la predicción para tres días posteriores y no más (Reyes, 2009).

Ahora bien, podemos realizar la predicción con el método estático con la restricción de que tan sólo nos proporcionará la estimación para la siguiente observación, en nuestro caso, la 286. Por lo que, si queremos estimar para la predicción 287 tenemos que ingresar el precio real de las acciones de ARA para la observación 286, es decir, necesitamos alimentar constantemente la serie. Sin embargo, este método es más preciso en virtud de que utiliza el valor real anterior en lugar de utilizar el valor o los valores estimados como el método dinámico.

Para llevar a cabo la predicción con el método estático estimamos la ecuación y oprimimos el botón “Forecast” en donde únicamente modificaremos el nombre de la serie por “ARAFF”; el método por el de estático (Static); y la muestra para la predicción por 1 a 286 (aunque el lector puede poner de 1 a 290 para que observe como tan sólo el programa predecirá para la observación 286)/Ok. Lo cual nos desplegará una gráfica como la siguiente y en el Workfile aparecerá la serie “araff”:

Grafico 11.4.11 Prediccion



Como puede observarse en la gráfica el precio estimado de las acciones de ARA (araff) con el método estático ésta más pegado a las bandas de confianza y ello se debe a que a través de este método de estimación, estático, se utilizan los valores reales y no los estimados. Por lo que la estimación con este método será más precisa. Asimismo, con respecto a las estadísticas, dado que casi todas (con la excepción de mean abs, percent error) tienen valores menores a uno, decimos que es buena estimación de la predicción.

El siguiente cuadro muestra el precio real, estimado y de predicción con el método estático para las observaciones 281 a la 286:

Tabla 11.4.4 Datos

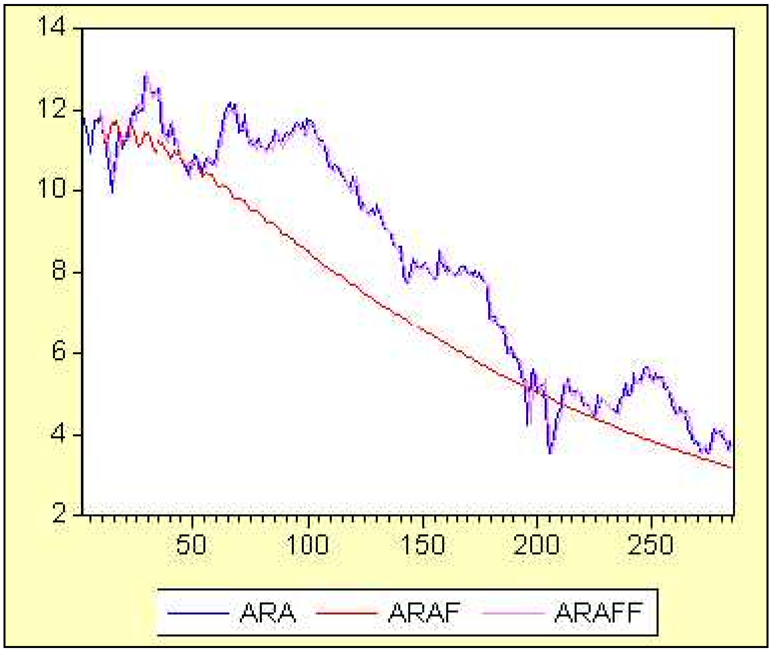
Obs.	Fecha	Precio real (Pr)	Precio estimado (Pe)	Error (Pr-Pe)
281	13/Feb/2009	3.87	3.98	-0.11
282	16/Feb/2009	3.86	3.83	0.03
283	17/Feb/2009	3.64	3.81	-0.17
284	18/Feb/2009	3.76	3.61	0.15
285	19/Feb/2009	3.86	3.77	0.09
286	20/Feb/2009	3.71	3.79	-0.08

Como se desprende del siguiente cuadro, el error de estimación (para el caso de las observaciones 281 a 285) y el error de predicción (para el caso de la observación 286) es muy pequeño. Por ejemplo, para el caso de la predicción del precio de las acciones del 20 de febrero de 2009 el error entre el precio real y el de predicción es de -0.08 pesos por acción, es decir, 8 centavos de peso. Error muy inferior al error del precio de predicción para la misma fecha con el método dinámico, a saber, 55 centavos de peso por acción.

En síntesis, para el caso de la predicción para la observación 286 (20 de febrero de 2009) el mejor método, es decir, el que se acerco más al precio real de las acciones de Consorcio ARA **es el estático**. Sin embargo, el método dinámico nos proporciona una **guía sobre el precio futuro** de las acciones y por tanto, decidir comprar o vender estas acciones.

Finalmente, el siguiente gráfico nos proporciona el precio real de las acciones de Consorcio ARA (línea de color azul), el precio estimado con el método dinámico (ARAF, línea de color rojo) y con el método estático (ARAFF, línea de color rosa), mismo que se obtiene seleccionando en el Workfile dichas series y dando click derecho sobre las mismas: Open/as Group/Ok., lo cual nos desplegará un cuadro con los valores de las series. Una vez ahí, vamos a: View/Graph/Line/Ok., operación que desplegará:

Grafico 11.4.12 Prediccion



Como puede observarse la predicción con el método estático (ARAFF) es más exacta que la predicción con el método dinámico (ARAF), ello debido a que, como se comento, el primero utiliza los valores reales para la estimación posterior y el segundo utiliza los valores estimados para las predicciones posteriores.

Ejercicio 2 sobre la aplicación de la metodología de Box & Jenkins

Con base en lo antes expuesto ahora procederemos a aplicarlo:

I.Identificación de la estacionariedad de la serie.

Detección de la estacionariedad

I.1.- Método informal.

En este ejercicio utilizamos el Tipo de Cambio nominal para México en el periodo de 1980 a 2010, utilizamos una base dividida en trimestres por lo que cada año cuenta con cuatro datos teniendo como resultado 124 datos por los 31 años comprendidos. Para ingresar la base en Eviews En la barra de herramientas elegimos File/New Worfile/ donde se abrirá un cuadro en el cual se elige la opción de datos regulares, ingresaremos del periodo de 1980.1 a 2010.4; nuestra base se encuentra en periodos trimestrales e incrementaremos 8 datos para la predicción de dos años.

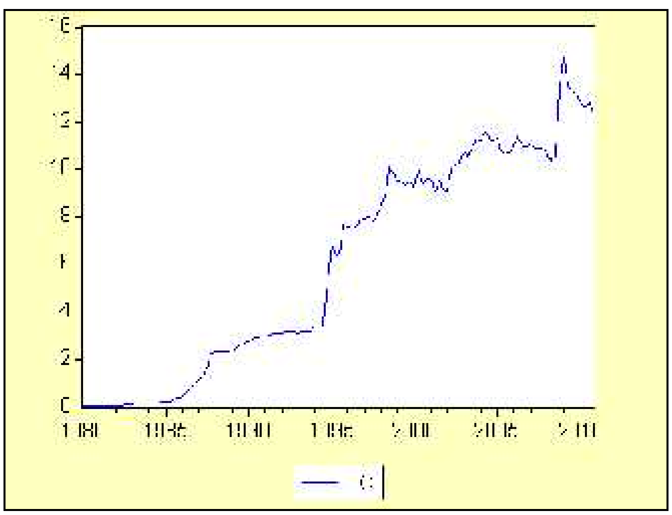
En el 678rea de Eviews de comandos procederemos a escribir data tc para generar la variable Tipo de Cambio, esto nos abrirá un cuadro donde fácilmente podemos pegar los datos.

Una vez cargada la serie cambiamos el periodo muestral (Sample) en la ventana del Workfile dando doble click sobre el mismo, es decir, sobre “Sample”. Este procedimiento desplegará una ventana denominada “Sample” donde escribiremos “ 1 124 para indicar que trabajaremos con los primeros 124 datos para estimar nuestros cálculos, sin embargo el programa sigue reconociendo los 130 espacios para usarlos posteriormente. ¿132?

c) Gráfica lineal y diagrama de dispersión.

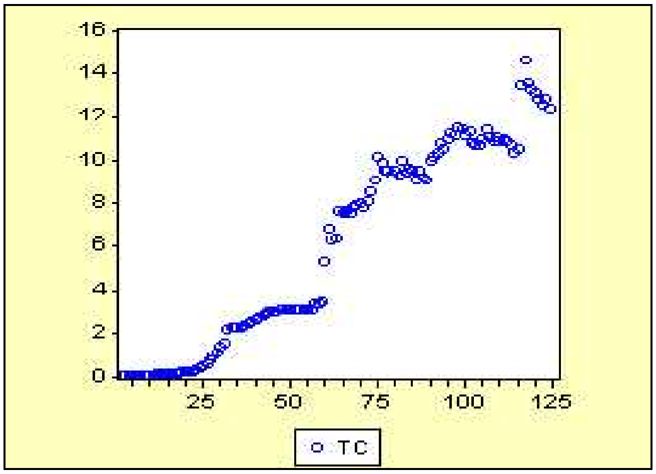
Ahora vamos a obtener la grafica de los datos de la variable Tipo de cambio, para esto damos doble click sobre la variable tc que abrirá los datos, posteriormente vamos a: View/Graph/Line/Ok. Lo que nos desplegará la siguiente grafica:

Grafico 11.4.13. Grafica Lineal “Tipo de Cambio”



Como puede observarse, la gráfica del proceso estocástico presenta un claro comportamiento creciente por lo que, podemos suponer que no es constante en media, por ello suponemos que el proceso no es estacionario. En el siguiente grafico de dispersión se observan datos muy dispersos por lo que tampoco es constante en varianza. Derivado de lo cual, quizá sea necesario y/o preciso trabajar en diferencias.

Grafico 11.4.14. Dispersion de Tipo de Cambio



d) Correlograma.

Otro procedimiento gráfico para comprobar la posible existencia de una raíz unitaria en el proceso estocástico consiste en inspeccionar el correlograma del mismo. Así, en la ventana de la serie vamos a: View/Correlogram…/, en donde se desplegará una ventana que nos solicita la especificación del correlograma donde indicaremos la opción Level (serie en niveles) y los rezagos que en predeterminado indica el programa que son 36 para después dar click en el botón de Aceptar que despliega el siguiente cuadro.

Cuadro 11.4.24. Correlograma

Sample: 1 126
Included observations: 124

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.980	0.980	122.03	0.000
		2	0.957	-0.006	239.39	0.000
		3	0.937	0.067	352.73	0.000
		4	0.915	-0.058	461.76	0.000
		5	0.892	-0.023	566.30	0.000
		6	0.869	-0.030	666.29	0.000
		7	0.845	-0.028	761.64	0.000
		8	0.818	0.092	851.72	0.000
		9	0.782	-0.227	934.83	0.000
		10	0.753	0.195	1012.6	0.000
		11	0.722	-0.129	1084.9	0.000
		12	0.690	-0.008	1151.3	0.000
		13	0.667	0.060	1211.9	0.000
		14	0.622	-0.053	1266.8	0.000
		15	0.589	0.046	1316.5	0.000
		16	0.550	-0.167	1360.2	0.000
		17	0.513	0.082	1398.7	0.000

Dado que como puede observarse en el correlograma de la serie tipo de cambio la función de autocorrelación desciende lentamente y la función de autocorrelación parcial presenta un valor significativo en el primer coeficiente, cercano a la unidad, debemos proceder a la transformación de la serie en logaritmos y posteriormente a diferenciarla. Asimismo, como todas las probabilidades asociadas al estadístico “Q” son menores a 0.05 se concluye que la serie es no estacionaria.

I.2.-Método formal.

Ahora pasamos a las pruebas formales para probar la existencia de raíz unitarua mediante Test de Dickey-Fuller (DF), una variante del mismo la Dickey-Fuller Aumentada (ADF) y el Test de Phillips-Perron (PP).

Prueba de hipótesis

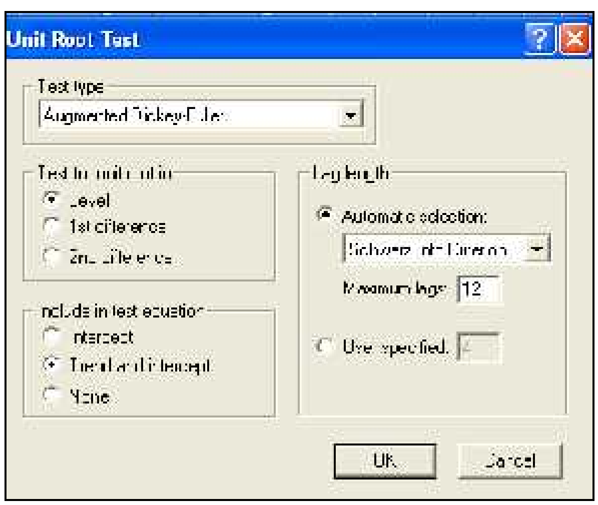
Ho: ρ = 1; existe raíz unitaria, por tanto, la serie es no estacionaria.

Ha: ρ < 1; no existe raíz unitaria, por tanto, la serie es estacionaria.

c) Test DF OJO: NO VIENE, ES ADF CON NIVELES

En Eviews para realizar la prueba DF una vez estando en la ventana de la serie vamos a: View/ UnitRoot Test…/ Dado que la serie original de tipo de cambio presenta tendencia creciente vamos a incluir en el Test el intercepto y el término de tendencia líneal como se muestra en la imagen 1 siguiente, después vamos a seleccionar los rezagos de forma manual o especificados por el usuario, en donde indicaremos que son cero (12?)para trabajar con el Test de DF, y; finalmente, vamos a realizar el Test primero en niveles, luego en primeras diferencias.

Imagen 11.4.15



Cuadro 11.4.25. Prueba DickeyFuller Aumentada

Null Hypothesis: TC has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.555195	0.3016
Test critical values:	1% level	-4.034356	
	5% level	-3.446765	
	10% level	-3.148399	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(TC)
Method: Least Squares
Date: 09/20/12 Time: 20:23
Sample (adjusted): 1980Q2 2010Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TC(-1)	-0.098429	0.038521	-2.555195	0.0119
C	-0.086629	0.106920	-0.810218	0.4194
@TREND(1980Q1)	0.012471	0.004950	2.519352	0.0131

R-squared	0.051753	Mean dependent var	0.100572
Adjusted R-squared	0.035949	S.D. dependent var	0.456907
S.E. of regression	0.448619	Akaike info criterion	1.258804
Sum squared resid	24.15113	Schwarz criterion	1.327394
Log likelihood	-74.41646	F-statistic	3.274638
Durbin-Watson stat	1.620450	Prob(F-statistic)	0.041238

Como el valor de t de ADF, 2.555 es menor que el valor t crítico de MacKinnon al 5%, en términos absolutos, con valor de 3.44 se acepta la Ho que nos dice que la serie presenta raíz unitaria por lo tanto la serie es no estacionaria. Asimismo, al ser la probabilidad asociada al estadístico DF mayor que 0.05 se acepta la Hipotesis nula.

Ahora, realicemos el Test de ADF pero en primeras diferencias con los demás criterios constantes, así obtenemos:

Cuadro 11.4.26 Prueba DickeyFuller 1era diferencia

Null Hypothesis: D(TC) has a unit root
Exogenous: Constant
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.368913	0.0000
Test critical values:	1% level	-3.484653	
	5% level	-2.885249	
	10% level	-2.579491	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(
Method: Least Squares
Sample (adjusted): 3 124
Included observations: 122 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(TC(-1))	-0.850091	0.090735	-9.368913	0.0000
C	0.085678	0.042354	2.022886	0.0453

R-squared	0.422456	Mean dependent var	-0.003443
Adjusted R-squared	0.417644	S.D. dependent var	0.597369
S.E. of regression	0.455866	Akaike info criterion	1.283020
Sum squared resid	24.93761	Schwarz criterion	1.328988
Log likelihood	-76.26422	F-statistic	87.77654
Durbin-Watson stat	1.908822	Prob(F-statistic)	0.000000

Dado que en términos absolutos el valor del estadístico ADF en primeras diferencias es, 9.36 mayor que el valor crítico de MacKinnon al 5% de significación igual a 2.88, se rechaza la Ho y se acepta la Ha, por tanto, no existe raíz unitaria, es decir, la serie es estacionaria. Asimismo, al ser la probabilidad asociada al estadístico DF menor a 0.05 se acepta la Ha y se rechaza la Ho.

En virtud de que el Test de ADF en primeras diferencias rechaza la Ho y acepta la Ha y dado que en niveles se acepta la Ho y se rechaza la Ha concluimos que la serie contiene una raíz unitaria y, por tanto, es integrada de orden 1, I(1). ¿ PARA QUE ESTA INTERPRETACION?

d) <u>Test ADF.</u>

Para elaborar este Test vamos a modificar únicamente el número de rezagos, así, en lugar de incluir cero como en el Test anterior de DF, vamos a incluir 1**.** Los demás términos los mantendremos constantes, de tal suerte que como la serie original de tipo de cambio presenta tendencia creciente vamos a incluir en el Test el intercepto y el término de tendencia lineal; luego, vamos a seleccionar los rezagos de forma manual o especificados por el usuario, en donde indicaremos que es uno para trabajar con el Test de ADF, y; finalmente, vamos a realizar el Test primero en niveles y luego en primeras diferencias para aceptar o rechazar la Ho. Así, siguiendo los pasos descritos, el resultado para el Test de ADF en niveles es:

Cuadro 11.4.27. Prueba ADF

Null Hypothesis: TC has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on AIC, MAXLAG=1)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.979338	0.1423
Test critical values:	1% level	-4.034997	
	5% level	-3.447072	
	10% level	-3.148578	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(TC)
Method: Least Squares
Date: 09/20/12 Time: 20:33
Sample (adjusted): 1980Q3 2010Q4
Included observations: 122 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TC(-1)	-0.116878	0.039230	-2.979338	0.0035
D(TC(-1))	0.202721	0.090040	2.251449	0.0262
C	-0.139942	0.110649	-1.264737	0.2085
@TREND(1980Q1)	0.014746	0.005056	2.916397	0.0042

R-squared	0.090678	Mean dependent var	0.101396
Adjusted R-squared	0.067559	S.D. dependent var	0.458699
S.E. of regression	0.442934	Akaike info criterion	1.241445
Sum squared resid	23.15047	Schwarz criterion	1.333380
Log likelihood	-71.72813	F-statistic	3.922325
Durbin-Watson stat	1.895039	Prob(F-statistic)	0.010383

Dado que el valor del estadístico ADF es menor en términos absolutos al valor crítico de MacKinnon, a saber -2.9793 < -3.4470 al 5% de significación estadística, y en virtud de que la probabilidad asociada al estadístico ADF es mayor a 0.05 se acepta la Ho y por tanto se dice que existe raíz unitaria, es decir, la serie es no estacionaria.

El resultado del Test de ADF con primeras diferencias manteniendo constante los demás términos es:

Cuadro 11.4.28. Prueba ADF con 1era diferencia.

Null Hypothesis: D(TC) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic based on SIC, MAXLAG=1)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.235894	0.0000
Test critical values:	1% level	-4.035648	
	5% level	-3.447383	
	10% level	-3.148761	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(TC,2)
Method: Least Squares
Date: 09/20/12 Time: 20:44
Sample (adjusted): 1980Q4 2010Q4
Included observations: 121 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(TC(-1))	-1.077899	0.116708	-9.235894	0.0000
D(TC(-1),2)	0.268322	0.089658	2.992713	0.0034
C	0.103097	0.083906	1.228724	0.2216
@TREND(1980Q1)	0.000108	0.001158	0.093165	0.9259

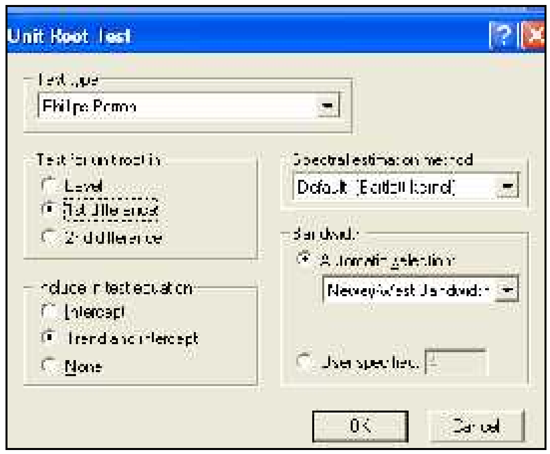
R-squared	0.463741	Mean dependent var	-0.003422
Adjusted R-squared	0.449991	S.D. dependent var	0.599342
S.E. of regression	0.444488	Akaike info criterion	1.248710
Sum squared resid	23.11561	Schwarz criterion	1.341133
Log likelihood	-71.54607	F-statistic	33.72604
Durbin-Watson stat	1.900602	Prob(F-statistic)	0.000000

Como puede observarse, al ser el valor del Test ADF mayor que el valor crítico de MacKinnon (9.23 > 3.44) en términos absolutos al 5% de significación estadística rechazamos la Ho; además, al ser la probabilidad asociada al Test de ADF menor a 0.05 (con un valor de 0.00) se corrobora el rechazo a la hipótesis nula y se concluye por lo tanto la variable tipo de cambio es estacionaria en primera diferencia.

e) **Test PP.**

Para obtener el Test Phillips Perron en Eviews es necesario abrir los datos tipo de cambio, ya estando en esta ventana vamos a View/Unit Root Test, y se abrirá un nuevo cuadro, para la prueba DF seleccionábamos este mismo en Test Type, ahora cambiaremos este tipo de prueba por la opción Pillips Perron como se muestra en la siguiente Imagen 2.

Imagen 11.4.16



El resultado del Test de PP en niveles, con intercepto y término de tendencia lineal y bandwidth predeterminado como 1 es:

Cuadro 11.4.29 Prueba PP

Phillips-Perron Unit Root Test on TC

Null Hypothesis: TC has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-2.742805	0.2216
Test critical values:	1% level	-4.034356	
	5% level	-3.446765	
	10% level	-3.148399	

*MacKinnon (1996) one-sided p-values.

Phillips-Perron Test Equation
Dependent Variable: D(TC)
Method: Least Squares
Date: 04/22/12 Time: 23:17
Sample (adjusted): 1980Q2 2010Q4
Included observations: 123 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TC(-1)	-0.098533	0.038557	-2.555479	0.0119
C	-0.086597	0.107004	-0.809287	0.4200
@TREND(1980Q1)	0.012481	0.004954	2.519066	0.0131

R-squared	0.051758	Mean dependent var	0.100569
Adjusted R-squared	0.035954	S.D. dependent var	0.457318
S.E. of regression	0.449022	Akaike info criterion	1.260596
Sum squared resid	24.19444	Schwarz criterion	1.329186
Log likelihood	-74.52666	F-statistic	3.275018
Durbin-Watson stat	1.620167	Prob(F-statistic)	0.041223

Como el valor del estadístico PP es, en términos absolutos, menor que el valor crítico de MacKinnon al 5% de significación estadística (-2.7428 < -3.4467) se acepta la Ho y dado que la probabilidad asociada al estadístico PP es mayor a 0.05 se acepta la Ho: existe raíz unitaria, por tanto, la serie es no estacionaria.

Ahora, realicemos la prueba PP en primeras diferencias manteniendo constante los demás términos de la selección, así obtenemos:

Cuadro 11.4.30 Prueba PP con 1era diferencia

Null Hypothesis: D(TC) has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 6 (Newey-West using Bartlett kernel)

		Adj. t Stat	Prob.*
Phillips-Perron test statistic		-9.131379	0.0000
Test critical values:	1% level	-4.034397	
	5% level	-3.447072	
	10% level	-3.148573	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.204032
HAC corrected variance (Bartlett kernel)	0.144355

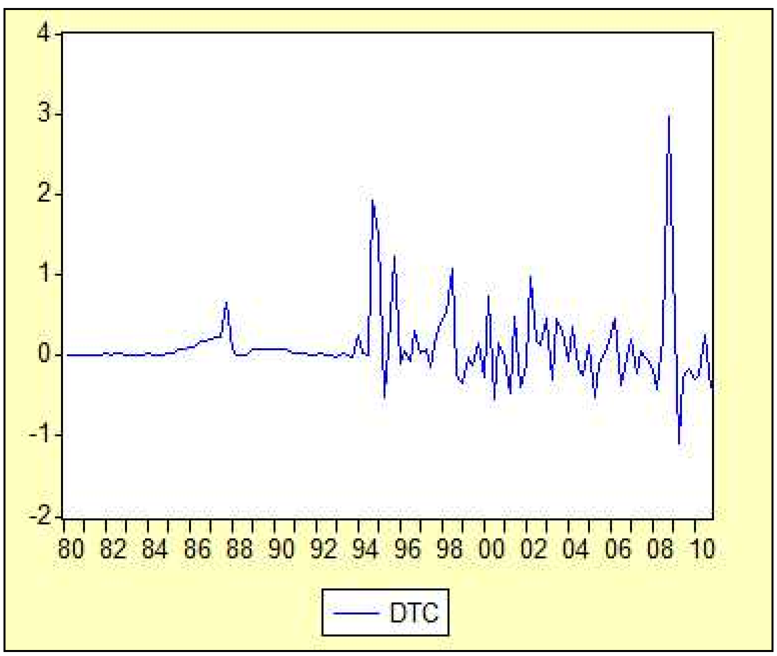
Al igual que en la prueba ADF en la prueba PP este estadístico presenta un valor mayor en relación los valores críticos de MacKinnon en términos absolutos(-9.181379 > -3.4470) y en virtud de que la probabilidad asociada al estadístico es menor que 0.05, por lo cual se rechaza H_0 : Tc presenta raíz unitaria, por lo tanto se acepta H_a : Tc no presenta raíz unitaria; por lo tanto la variable tipo de cambio es estacionaria en primera diferencia**.**

II) <u>Transformación y diferenciación de la serie.</u>

Se procederá a transformar la serie de tipo de cambio en diferencias, en el ejemplo anterior se utilizaron logaritmos para reducir la dispersión de los datos. En este caso el tipo de cambio está expresando en la proporción que existe entre el valor de una moneda con otra. En eviews para obtener una diferencia en el area de comandos generamos la variable escribiendo genr dtc=d(tc) el algoritmo en eviews para diferenciar es solamente "D".¿ PARA QUE D?

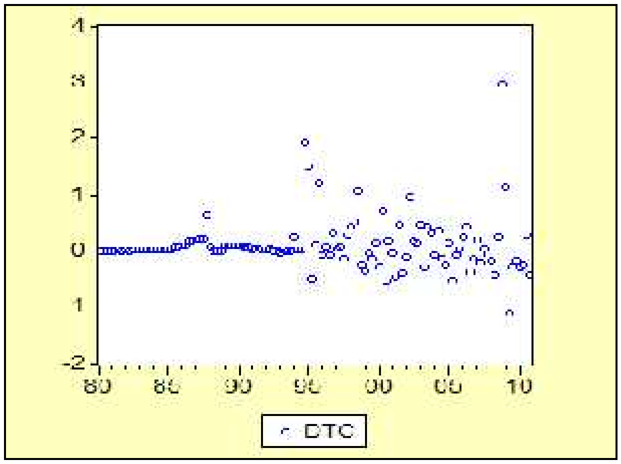
A continuación procedemos a graficarla de forma lineal a través de los pasos ya descritos anteriormente. De esta forma obtenemos

Grafico 11.4.15. DTC



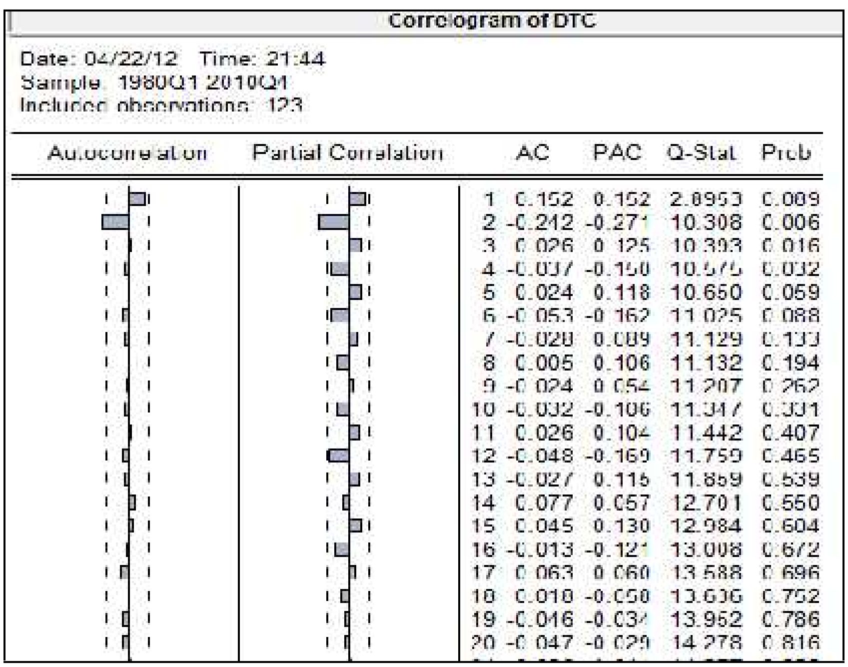
Como puede apreciarse, la serie "DTC" presenta gráficamente un comportamiento paralelo al eje de las abscisas, desde inicios de la década de los ochenta hasta mediados de la década de los noventa presentan un comportamiento estable, para los siguiente años son más notables las fluctuaciones en las diferencias del tipo de cambio. Esto puede indicar una media constante. Asimismo, el diagrama de dispersión nos muestra que la mayoría de las variables aleatorias oscilan alrededor de la media en el periodo anteriormente descrito, a partir de mediados de la década de los noventa los datos se encuentran más dispersos, puede observarse en la siguiente gráfica.

Gráfica 11.4.16. Diagrama de dispersión de las diferencias del Tipo de Cambio.



Procedemos ahora realizar el correlograma de la serie "DTC" para verificar la estacionariedad de la misma. Así, con los pasos ya conocidos obtenemos el siguiente correlograma:

Cuadro 11.4.31. Correlograma de DTC.



Correlogram of DTC

Date: 04/22/12 Time: 21:44
Sample: 1980Q1 2010Q4
Included observations: 123

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.152	0.152	2.8953	0.089
		2	-0.212	-0.271	10.308	0.006
		3	0.026	0.125	10.393	0.016
		4	-0.037	-0.150	10.575	0.032
		5	0.024	0.118	10.650	0.059
		6	-0.053	-0.162	11.025	0.088
		7	-0.028	0.089	11.129	0.133
		8	0.005	0.106	11.132	0.194
		9	-0.024	0.054	11.207	0.262
		10	-0.032	-0.106	11.347	0.331
		11	0.026	0.104	11.442	0.407
		12	-0.048	-0.169	11.759	0.465
		13	-0.027	0.115	11.859	0.539
		14	0.077	0.057	12.701	0.550
		15	0.045	0.130	12.984	0.604
		16	-0.013	-0.121	13.008	0.672
		17	0.063	0.060	13.588	0.696
		18	0.018	-0.050	13.636	0.752
		19	-0.046	-0.034	13.952	0.786
		20	-0.047	-0.029	14.278	0.816

Como puede observarse, al encontrarse los valores de la autocorrelación (AC) y los de la autocorrelación parcial (PAC) dentro de las bandas de confianza, al ser los valores de AC y PAC muy menores a 1 y al ser la probabilidad asociada al estadístico “Q” mayores a 0.05, con excepción del segundo periodo que la probabilidad es menor a 0.05. Podemos concluir, de acuerdo a este correlograma, que la serie “DTC” probablemente es estacionaria, es decir, no presenta raíz unitaria.

Ahora, realicemos los Test de raíces unitarias correspondientes, es decir, las pruebas o métodos formales, para determinar que la serie efectivamente es ya estacionaria.

d) <u>Test de ADF para “DTC”.</u>

Siguiendo los pasos ya descritos con antelación para realizar la prueba DF y, planteando la prueba de hipótesis correspondiente de la siguiente manera:

Ho: $\rho = 1$; existe raíz unitaria, por tanto, la serie es no estacionaria.

Ha: $\rho < 1$; no existe raíz unitaria, por tanto, la serie es estacionaria.

Cabe hacer mención que para realizar la prueba ADF va a hacer en niveles (Level) porque ya está diferenciada una vez, por lo que si se selecciona primeras diferencias en realidad se tratará de las segundas diferencias; dado que la nueva serie (DTC) no parece presentar tendencia (creciente o decreciente) pero parece tener uno o más valores medios distintos de cero debemos incluir un término constante (intercept) en la regresión; y en el número de rezagos cero para realizar el Test DF. El resultado del Test de DF es:

Cuadro 11.4.32. Prueba ADF para DTC

Augmented Dickey-Fuller Unit Root Test on DTC

Null Hypothesis: DTC has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.368913	0.0000
Test critical values:	1% level	-3.484653	
	5% level	-2.885249	
	10% level	-2.579491	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DTC)
Method: Least Squares
Date: 04/22/12 Time: 21:59
Sample (adjusted): 1980Q3 2010Q4
Included observations: 122 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DTC(-1)	-0.850091	0.090735	-9.368913	0.0000
C	0.085678	0.042354	2.022886	0.0453

R-squared	0.422456	Mean dependent var	-0.003443
Adjusted R-squared	0.417644	S.D. dependent var	0.597369
S.E. of regression	0.455866	Akaike info criterion	1.283020
Sum squared resid	24.93761	Schwarz criterion	1.328988
Log likelihood	-76.26422	F-statistic	87.77654
Durbin-Watson stat	1.908822	Prob(F-statistic)	0.000000

Como el valor del estadístico DF (-9.3689) en términos absolutos es mayor que el valor crítico de MacKinnon al 5% (-2.8852) y dado que la probabilidad del estadístico DF es menor a 0.05 se rechaza la Ho, por tanto, no existe raíz unitaria, es decir, la serie “DTC” es estacionaria.

e) Test de ADF para “DTC”.

Ahora realicemos la prueba ADF, manteniendo los términos seleccionados constantes pero modificando el número de rezagos a 1 en lugar de cero. Obtenemos:

Cuadro 11.4.33. Prueba ADF para DTC con 1era diferencia

Augmented Dickey-Fuller Unit Root Test on DTC

Null Hypothesis: DTC has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic based on SIC, MAXLAG=1)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.285192	0.0000
Test critical values:	1% level	-3.485115	
	5% level	-2.885450	
	10% level	-2.579598	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DTC)
Method: Least Squares
Date: 04/22/12 Time: 22:01
Sample (adjusted): 1980Q4 2010Q4
Included observations: 121 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DTC(-1)	-1.078311	0.116132	-9.285192	0.0000
D(DTC(-1))	0.269087	0.089252	3.014919	0.0031
C	0.109931	0.042058	2.613821	0.0101

R-squared	0.463923	Mean dependent var	-0.003471
Adjusted R-squared	0.454837	S.D. dependent var	0.599852
S.E. of regression	0.442902	Akaike info criterion	1.233545
Sum squared resid	23.14714	Schwarz criterion	1.302862
Log likelihood	-71.62945	F-statistic	51.05876
Durbin-Watson stat	1.908162	Prob(F-statistic)	0.000000

En virtud de que el valor del estadístico ADF (-9-2851) es mayor que el valor crítico de MacKinnon al 5% (-2.8854) y dado que la probabilidad asociada al estadístico ADF es menor a 0.05 se rechaza Ho, por tanto, no existe raíz unitaria, es decir, la serie “DTC” es estacionaria.

f) <u>Test de PP para “DTC”.</u>

Para el caso del Test de PP vamos a realizarlo en niveles (Level); con intercepto (intercept); y vamos a especificar de forma manual 1 rezago. Así obtenemos lo siguiente:

Cuadro 11.4.34 Prueba PP para DTC.

Phillips-Perron Unit Root Test on DTC

Null Hypothesis: DTC has a unit root
Exogenous: Constant
Bandwidth: 6 (Newey-West using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-9.229597	0.0000
Test critical values:	1% level	-3.484653	
	5% level	-2.885249	
	10% level	-2.579491	

*MacKinnon (1996) one-sided p-values.

Phillips-Perron Test Equation
Dependent Variable: D(DTC)
Method: Least Squares
Date: 04/22/12 Time: 22:07
Sample (adjusted): 1980Q3 2010Q4
Included observations: 122 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DTC(-1)	-0.850091	0.090735	-9.368913	0.0000
C	0.085678	0.042354	2.022886	0.0453

R-squared	0.422456	Mean dependent var	-0.003443
Adjusted R-squared	0.417644	S.D. dependent var	0.597369
S.E. of regression	0.455866	Akaike info criterion	1.283020
Sum squared resid	24.93761	Schwarz criterion	1.328988
Log likelihood	-76.26422	F-statistic	87.77654
Durbin-Watson stat	1.908822	Prob(F-statistic)	0.000000

Como puede observarse en los resultados del Test de PP el estadístico PP es mayor (-9.2295) al valor crítico de MacKinnon al 5% (-2.8852) y dado que la probabilidad asociada al estadístico PP es menor a 0.05, entonces rechazamos la Ho, por lo que no existe raíz unitaria, luego entonces la serie “DTC” es estacionaria.

Dado que con los métodos informales (gráfica lineal, diagrama de dispersión y correlograma) concluimos que la serie “dtc”, probablemente es estacionaria; y con los métodos formales (Test ADF y PP) se acepta la Ha en virtud de que los valores de los estadísticos son mayores a los valores críticos de MacKinnon al 5% y a que la probabilidad asociada a los mismos es menor a 0.05, concluimos c que la serie de tiempo “DTC” es estacionaria. Por tanto, podemos pasar a la identificación del modelo ARMA, pues ya sabemos que la serie es integrada de orden 1, es decir, I(1).

Otra prueba que es muy utilizada es la KPSS a diferencia de las anteriores pruebas donde la hipótesis nula es la existencia de raíz unitaria, en la KPSS la hipótesis nula es la estacionariedad de la serie.

La regla de decisión es la siguiente:

H_o : la serie es estacionaria.

H_a: la serie es no estacionaria.

Estadístico KPSS > al valor de tabla (1, 5 y 10%) se rechaza H_o

Nota: a diferencia de las pruebas ADF, DF y PP donde el valor del estadístico en términos absolutos tenía que ser mayor a los valores de MacKinnon para rechazar H_o y aceptar la H_a para que la serie no presentara raíz unitaria y por lo tanto deducir que la serie es estacionaria, en la KPSS si el valor del estadístico es mayor a los valores de tablas se rechaza H_0 por lo tanto la serie es no estacionaria, *en resumen se busca aceptar la hipótesis nula de estacionariedad en la serie.*

I.3.-Identificación de la estructura ARMA.

Para identificar los procesos que se incluirán en el modelo o los modelos posibles a estimar, es necesario contar con las funciones de autocorrelación y autocorrelación parcial del correlograma de la serie estacionaria, en nuestro caso de la serie: DTC. Como puede observarse en el correlograma, el segundo rezago sale de las bandas de confianza tanto de las funciones de autocorrelación como de autocorrelación parcial.

De tal suerte y derivado del análisis visual realizado a través del correlograma para el caso de la serie DTC y tras realizar diversos modelos con ayuda de Eviews, los modelos identificados que vamos a desarrollar aquí son:

3. AR(2) MA(2), es decir, ARMA(2,2) y en virtud de que diferenciamos una vez la serie para convertirla en estacionaria es integrada de orden 1, I(1). Por lo tanto el modelo que se desarrolla es un ARIMA(2,1,2).

I.3.1.-Estimación de coeficientes y evaluación-validación del modelo.

Así, una vez identificado el modelo o los modelos procedemos a estimarlos con ayuda de Eviews 5 como se había descrito en el ejercicio anterior.

Cuadro 11.4.35. Estimación modelo ARMA.

Dependent Variable: DTC
Method: Least Squares
Date: 04/22/12 Time: 22:21
Sample (adjusted): 1980Q4 2010Q4
Included observations: 121 after adjustments
Convergence achieved after 14 iterations
Backcast: 1980Q2 1980Q3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.115494	0.018838	6.130901	0.0000
AR(2)	0.550983	0.152741	3.607306	0.0005
MA(2)	-0.808000	0.107317	-7.529114	0.0000

R-squared	0.090648	Mean dependent var	0.102231
Adjusted R-squared	0.075236	S.D. dependent var	0.460927
S.E. of regression	0.443249	Akaike info criterion	1.235113
Sum squared resid	23.18346	Schwarz criterion	1.304430
Log likelihood	-71.72431	F-statistic	5.881389
Durbin-Watson stat	1.561528	Prob(F-statistic)	0.003674

Inverted AR Roots	.74	-.74
Inverted MA Roots	.90	-.90

Con los resultados obtenidos hay que juzgar la validez del modelo estimado, para ello debemos verificar:

- Contraste "t" de significación de coeficientes: Para nuestro caso, al ser la probabilidad asociada a las pruebas "t" para cada coeficiente menores que 0.05 concluimos que los coeficientes son estadísticamente significativos.

- Matriz de covarianzas. Se obtienen con los pasos descritos en el ejercicio anterior:

Cuadro 11.4.35. Matriz de covarianzas.

	C	AR(2)	MA(2)
C	0.000355	0.000318	-0.000255
AR(2)	0.000318	0.023330	-0.014146
MA(2)	-0.000255	-0.014146	0.011517

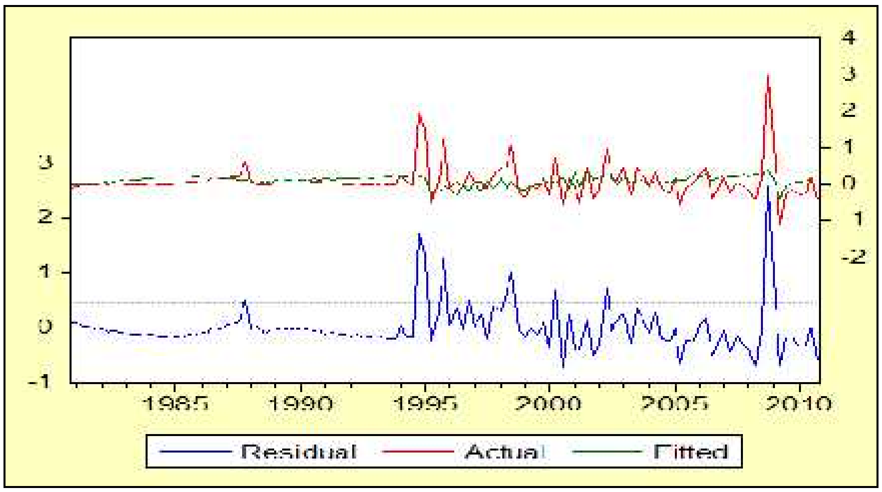
En este caso, como se recordara, lo que nos interesa analizar son los valores fuera de la diagonal principal.

Para nuestro caso, observamos que la covarianza que existe entre los procesos AR(2) MA(2) y entre cada uno de estos con la constante es muy reducida, por tanto, concluimos que debemos mantener todos los coeficientes del modelo y podemos seguir adelante

- <u>Tabla y gráfico de residuos.</u> Para acceder a la tabla que nos muestra los residuos, así como los valores de la serie original y estimada, dentro del menú de la ventana de la ecuación del modelo estimado seleccionamos.

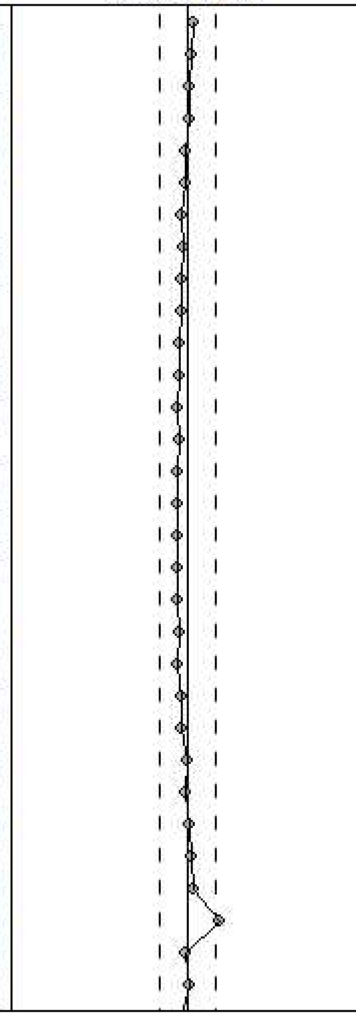
Para nuestro caso la gráfica de los residuos obtenidos a través del modelo planteado es

Grafico 11.4.17. Residuos



Cuadro 11.4.36. Tabla de residuos.

obs	Actual	Fitted	Residual	Residual Plot
1980Q4	0.00000	-0.08154	0.08154	
1981Q1	0.00000	-0.07458	0.07458	
1981Q2	0.00000	-0.01402	0.01402	
1981Q3	0.01000	-0.00840	0.01840	
1981Q4	0.00000	0.04053	-0.04053	
1982Q1	0.02000	0.04250	-0.02250	
1982Q2	0.00000	0.08461	-0.08461	
1982Q3	0.02000	0.08106	-0.06106	
1982Q4	0.03000	0.12022	-0.09022	
1983Q1	0.01000	0.11221	-0.10221	
1983Q2	0.01000	0.14129	-0.13129	
1983Q3	0.01000	0.13996	-0.12996	
1983Q4	0.01000	0.16345	-0.15345	
1984Q1	0.02000	0.16237	-0.14237	
1984Q2	0.01000	0.18135	-0.17135	
1984Q3	0.01000	0.17792	-0.16792	
1984Q4	0.01000	0.19582	-0.18582	
1985Q1	0.02000	0.19305	-0.17305	
1985Q2	0.02000	0.20751	-0.18751	
1985Q3	0.07000	0.20270	-0.13270	
1985Q4	0.07000	0.21439	-0.14439	
1986Q1	0.10000	0.19765	-0.09765	
1986Q2	0.10000	0.20709	-0.10709	
1986Q3	0.18000	0.18586	-0.00586	
1986Q4	0.17000	0.19349	-0.02349	
1987Q1	0.20000	0.15577	0.04423	
1987Q2	0.23000	0.16451	0.06549	
1987Q3	0.22000	0.12632	0.09368	
1987Q4	0.64000	0.12567	0.51433	
1988Q1	0.07000	0.09738	-0.02738	
1988Q2	0.00000	-0.01109	0.01109	



Como puede observarse en la gráfica de los valores reales, estimados y residuales, estos últimos se muestran dentro, durante la mayor parte del periodo, de las bandas de confianza, con excepción del periodo comprendido a mediados de la década de los noventa, periodo de gran volatilidad del tipo de cambio debido al error de diciembre y porque se establece que el tipo de cambio debe ser flotante.

- Prueba F global: Como se sabe la prueba F es un contraste de significación de conjunto de todo el modelo. Así, si la probabilidad asociada al estadístico F es menor que 0.05 se concluye que las variables independientes en su conjunto explican el comportamiento de la variable dependiente y viceversa. En nuestro caso, al ser la probabilidad asociada a F menor que 0.05 concluimos que el modelo en su conjunto es adecuado.

- <u>Coeficiente de determinación (R^2):</u> En nuestro caso, tenemos que $R^2 = 0.090648$.

- <u>Correlograma de los residuos:</u> Siguiendo con la validación del modelo, vamos a observar el correlograma de los residuos obtenidos a partir del modelo planteado.

- <u>Test DF, ADF y PP para los residuos:</u> Para verificar que los residuos son ruido blanco, podemos ahora realizar sobre ellos los Test DF, ADF Y PP.

Cuadro 11.4.37. Correlograma u1.

Correlogram of U1

Date: 04/22/12 Time: 22:42
Sample: 1980Q1 2010Q4
Included observations: 122

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.132	0.132	2.1733	0.140
		2	0.178	0.163	6.1686	0.046
		3	0.092	0.053	7.2430	0.065
		4	0.054	0.010	7.6166	0.107
		5	0.014	-0.017	7.6430	0.177
		6	-0.100	-0.121	8.9524	0.176
		7	-0.135	-0.124	11.332	0.125
		8	-0.075	-0.018	12.087	0.147
		9	-0.116	-0.054	13.891	0.126
		10	-0.072	-0.015	14.596	0.147
		11	0.086	0.035	15.613	0.156
		12	0.030	0.003	15.734	0.204
		13	-0.076	-0.073	16.544	0.221
		14	0.084	0.098	17.538	0.229
		15	0.024	0.013	17.618	0.283
		16	0.055	0.011	18.049	0.321
		17	-0.076	-0.131	18.091	0.335
		18	0.017	-0.007	18.933	0.396
		19	0.117	0.139	20.929	0.341
		20	-0.013	0.011	20.956	0.400
		21	-0.080	-0.027	21.907	0.405
		22	-0.145	-0.121	25.083	0.293
		23	-0.087	-0.053	26.243	0.290
		24	-0.107	-0.160	31.621	0.137
		25	-0.126	-0.094	34.111	0.106
		26	-0.153	-0.145	37.809	0.063
		27	0.149	0.107	41.341	0.038
		28	0.013	0.001	41.368	0.050
		29	0.058	0.055	41.912	0.057
		30	0.098	0.005	43.505	0.053

Pruebas de Raíz Unitaria a U1

Cuadro 11.4.38. ADF con cero rezagos.

Augmented Dickey-Fuller Unit Root Test on U1

Null Hypothesis: U1 has a unit root
Exogenous: None
Lag Length: 0 (Automatic based on SIC, MAXLAG=0)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.627742	0.0000
Test critical values:	1% level	-2.584214	
	5% level	-1.943494	
	10% level	-1.614970	

*MacKinnon (1996) one-sided p-values.

Cuadro 11.4.39. ADF con 1 rezago.

Augmented Dickey-Fuller Unit Root Test on U1

Null Hypothesis: U1 has a unit root
Exogenous: None
Lag Length: 0 (Automatic based on SIC, MAXLAG=1)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.627742	0.0000
Test critical values:	1% level	-2.584214	
	5% level	-1.943494	
	10% level	-1.614970	

*MacKinnon (1996) one-sided p-values.

Cuadro 11.4.40. PP con 1 rezago.

Phillips-Perron Unit Root Test on DTC

Null Hypothesis: DTC has a unit root
Exogenous: None
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-9.074151	0.0000
Test critical values:	1% level	-2.584055	
	5% level	-1.943471	
	10% level	-1.614984	

*MacKinnon (1996) one-sided p-values.

Como en las tres pruebas la probabilidad asociada a cada estadístico es menor a 0.05 y en virtud de que el valor de los estadísticos DF, ADF y PP son mayores a los valores críticos de MacKinnon al 5% concluimos que la serie de los residuos no presenta raíz unitaria, por tanto, es ruido blanco, es decir, tiene un comportamiento aleatorio.

Ahora realicemos el segundo modelo planteado siguiendo y aplicando los mismos pasos y criterios descritos en el punto IV para el primer modelo.

Cuadro11.4.41. Restimación del modelo SARIMA.

Dependent Variable: DTC
Method: Least Squares
Date: 09/20/12 Time: 21:20
Sample (adjusted): 1980Q4 2010Q4
Included observations: 121 after adjustments
Convergence achieved after 0 iterations
Backcast: 1980Q1 1980Q3

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.109926	0.026133	4.205605	0.0001
AR(2)	0.371933	0.199750	1.861996	0.0451
MA(2)	-0.686024	0.154441	-4.441972	0.0000
SMA(1)	0.279679	0.092331	3.029081	0.0030

R-squared	0.146734	Mean dependent var	0.102233
Adjusted R-squared	0.120029	S.D. dependent var	0.460510
S.E. of regression	0.431059	Akaike info criterion	1.187355
Sum squared resid	21.73997	Schwarz criterion	1.279770
Log likelihood	-67.83497	F-statistic	6.653203
Durbin-Watson stat	2.000576	Prob(F-statistic)	0.000343

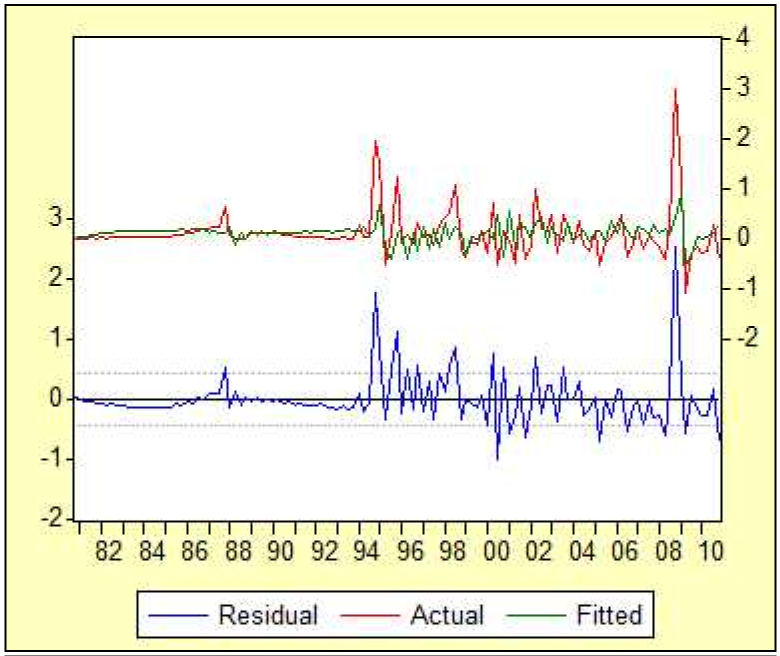
- **<u>Matriz de covarianzas:</u>**

Cuadro 11.4.42. Matriz de covarianzas.

	C	AR(2)	MA(2)	SMA(1)
C	0.000671	0.000280	-0.000261	-2.40E-05
AR(2)	0.000280	0.038864	-0.026645	0.002773
MA(2)	-0.000261	-0.026645	0.022949	-0.000823
SMA(1)	-2.40E-05	0.002773	-0.000823	0.008519

- <u>Tabla y gráfica de los residuos:</u>

Gráfica 1.4.18. Gráfico de los residuos.



Cuadro 11.4.43. Tabla de residuos.

obs	Actual	Fitted	Residual	Residual Plot
1980Q4	0.00000	-0.02297	0.02297	
1981Q1	0.00000	0.01689	-0.01689	
1981Q2	0.00000	0.03955	-0.03955	
1981Q3	0.01000	0.06489	-0.05489	
1981Q4	0.00000	0.08399	-0.08399	
1982Q1	0.02000	0.09453	-0.07453	
1982Q2	0.00000	0.11655	-0.11655	
1982Q3	0.02000	0.11131	-0.09131	
1982Q4	0.03000	0.13814	-0.10814	
1983Q1	0.01000	0.13162	-0.12162	
1983Q2	0.01000	0.13826	-0.12826	
1983Q3	0.01000	0.14150	-0.13150	
1983Q4	0.01000	0.14780	-0.13780	
1984Q1	0.02000	0.14955	-0.12955	

- <u>Correlograma de los residuos:</u>

Cuadro 11.4.44. Correlograma U2.

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.0052	
		2	-0.016	-0.017	0.0393	
		3	0.115	0.115	1.7140	
		4	0.009	0.010	1.7236	0.189
		5	0.056	0.061	2.1322	0.344
		6	-0.004	-0.017	2.1344	0.545
		7	-0.022	-0.022	2.1964	0.700
		8	0.056	0.043	2.6116	0.760
		9	-0.030	-0.029	2.7282	0.842
		10	-0.002	0.002	2.7289	0.909
		11	0.023	0.013	2.8026	0.946
		12	-0.027	-0.019	2.9036	0.968
		13	-0.055	-0.061	3.3163	0.973
		14	0.079	0.079	4.1905	0.964
		15	-0.020	-0.016	4.2486	0.979
		16	0.018	0.030	4.2970	0.988
		17	-0.098	-0.113	5.6583	0.974

Comentario sobre el correlograma: Como puede observarse en las funciones de autocorrelación (AC) y autocorrelación parcial (PAC) ninguna sale de las bandas de confianza, asimismo, los valores de AC y PAC son muy menores a 1; y la probabilidad asociada al estadístico Q son mayores que 0.05. Derivado de estos resultados podemos concluir que los residuos son ruido blanco.

- Test DF, ADF y PP para los residuos (en ese orden):

Cuadro 11.4.45. Prueba Dickey-Fuller U2.

Augmented Dickey-Fuller Unit Root Test on D(U2)

Null Hypothesis: D(U2) has a unit root
Exogenous: None
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-18.60253	0.0000
Test critical values:	1% level	-2.584707	
	5% level	-1.943563	
	10% level	-1.614927	

*MacKinnon (1996) one-sided p-values.

Cuadro 11.4.46. Prueba Ampliada de Dickey-Fuller U2.

Augmented Dickey-Fuller Unit Root Test on D(U2)

Null Hypothesis: D(U2) has a unit root
Exogenous: None
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-14.51067	0.0000
Test critical values:	1% level	-2.584877	
	5% level	-1.943587	
	10% level	-1.614912	

*MacKinnon (1996) one-sided p-values.

Cuadro 11.4.47. Prueba Phillips-Perron U2.

Phillips-Perron Unit Root Test on D(U2)

Null Hypothesis: D(U2) has a unit root
Exogenous: None
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-19.79953	0.0000
Test critical values:	1% level	-2.584707	
	5% level	-1.943563	
	10% level	-1.614927	

*MacKinnon (1996) one-sided p-values.

Para elegir entre un modelo y otro se consideran los planteamientos del ejercicio anterior, lo cuales son: la suma de cuadrados de los residuos (Sum squaredresid) y el error estándar de la regresión (S. E. of regresión). Se selecciona el modelo que presente la suma de los errores al cuadrado y el error estándar de la regresión menores. Así mismo, a pesar del problema de utilizar el R^2 como una prueba de la bondad de ajuste del modelo, dado que éste se ve fuertemente influenciado por las transformaciones de la serie, cuando se trata de discernir entre varios modelos nos quedaremos con aquel que posea el R^2 más alto.

A continuación se presentan los resultados obtenidos en los dos modelos:

Criterios	Primer modelo: ARIMA 2,1, 2)	Segundo modelo: SARMA (2,2,1)
Suma de cuadrado de los residuos	23.18349	21.73051
Desviación estándar de la regresión	0.443249	0.430965
Coeficiente de determinación ajustado	0.075236	0.125784
Logaritmo de verosimilitud (L)	-71.72431	-67.80864
Criterio de información de Akaike (AIC)	1.235113	1.186920
Criterio de Schwarz (SC)	1.304430	1.279342

FALTO LA SELECCIÓN Y EL CIERRE CON LA PREDICCION

CAPITULO XII.- MODELOS DINAMICOS: DE VECTORES AUTORREGRESIVOS (VAR) Y VECTORES DE ERRORES CORREGIDOS (VEC), COINTEGRACION y PRONÓSTICO.

12.1. Los modelos VAR: VECTORES AUTORREGRESIVOS.

Son una generalización al campo multivariante de los modelos autorregresivos univariantes AR de Box y Jenkins. En sus versiones modernas (modelos VARMA) tienen un planteamiento similar a los modelos multvariantes de funciones de transferencia (Pérez, 2006).

Gujarati (2003:823) señala que se parecen a los modelos de ecuaciones simultáneas porque trabajan varias variables endógenas a la vez, donde cada una de ellas es explicada por sus valores pasados o rezagados y por los valores rezagados del resto de las variables que componen el modelo.

Su utilidad se debe a que sirven para hacer predicciones. Sus antecedentes se localizan en modelos desarrollados con el Sistema de Ecuaciones Simultáneas, **SES**, como también en la recomendación de su precursor **Christopher Sims** de que para que haya verdadera simultaneidad entre las variables a todas se les debe de tratar por igual, y que por ende no debe existir ninguna distinción apriorística entre las variables endógenas del modelo. Sus recomendaciones también se derivan del desarrollo previo que hizo Granger de la prueba de causalidad, en la que *no hay variables exógenas en este sistema de ecuaciones.*

Al respecto, debe indicarse que el término autorregresivo aquí se refiere al uso del valor rezagado de la variable dependiente en el lado derecho de la ecuación; **igualmente, el término vector es indicativo de que se están manejando dos o más variables cuyos valores se autodeterminan simultáneamente: influencia recíproca. (Gujarati, 2004: 823)**

Un caso típico de estos modelos es cuando se estima la influencia recíproca entre las variables, es decir, la causalidad de una sobre la otra y viceversa; en este caso las U_t también como en ejemplos anteriores son los términos de error en cada ecuación del **SES**, pero aquí se les denomina *impulso, innovaciones o choques pasados* (Pérez, 2006:503).

Para estimar el valor de los estimadores del **SES** con p*uras variables endógenas,* que aparecen como “exógenas” en el lado derecho de las ecuaciones con sólo valores rezagados, **antes** es necesario *determinar la longitud del rezago K,* determinación que generalmente es empírica , sabiendo de antemano el experto que usar *muchos* valores rezagados consume demasiados grados de libertad y es posible que surja la multicolinealidad y, que, cuando se incluyen *pocos* rezagos puede incurrirse en errores de especificación.

En ese sentido la opción técnica para determinar la longitud del rezago es utilizar los criterios de AKAIKE y de SCHWARZ y seleccionar entre ellos dos el *de menor valor*; acompañado de un alto valor del coeficiente de determinación. Como se intuye, en cierta forma la determinación de la longitud del rezago k es una prueba de ensayo y error. Algunos estudiosos sugieren que se use un número de rezagos equivalente a un tercio del total de observaciones que tiene la serie temporal.

Cuando se usa Eviews para calcular con MCO los parámetros del **SES**, que con propósitos ilustrativos digamos que corresponden a la ecuación de la oferta monetaria (M_t) y la tasa de interés (R_t), que cuando actúan como "exógenas" en el lado derecho de cada una de las dos ecuaciones, *sus valores correspondientes son términos rezagados*s. Así, al especificar las variables oferta monetaria (M_t) y tasa de interés (R_t), cada rezago se especifica así: M_{t-1} , M_{t-2}, …….., M_{t-k}) y R_{t-1}, R_{t-2},……….., R_{t-k} .

Lo anterior visto como un proceso autorregresivo (AR) en opinión de Gujarati (2004: 825) se representa así:

Sea la serie temporal Y_t del PIB en el periodo t, de manera que si la modelamos como

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + U_t;$$

donde δ es la media de Y_t y U_t es un término de error aleatorio no correlacionado con media igual a cero y varianza constante,i.e., ruido blanco. En este caso se dice que Y_t sigue un proceso estocástico regresivo de primer orden o AR(1). Aquí el valor de Y_t en el momento t depende de su valor en el periodo anterior (Y_{t-1}) y de un término aleatorio U_t. También se observa que los valores de Y están expresados como desviaciones de su media aritmética δ. Comenta este autor que este modelo indica que el valor de prónostico de Y en el periodo t es tan sólo alguna proporción α_1 de su valor en el periodo t_{-1} más " un choque" o perturbación en el tiempo t.

Ahora bien, si el modelo se plantea así:

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \alpha_2 (Y_{t-2} - \delta) + U_t$$

En este caso se dice que Y_t sigue un proceso estócástico regresivo de segundo orden o AR(2). Esto en otras palabras indica que el valor de Y_t en el periodo t depende de sus valores en los dos periodos anteriores más U_t.

Señala que en general

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \alpha_2 (Y_{t-2} - \delta) + \ldots\ldots \alpha_k (Y_{t-k} - \delta) + U_t$$

En donde Y_t es un proceso estocástico autorregresivo de orden k o AR(k).

Se menciona entre los especialistas en este tema que al manejar en estos modelos solamente los valores actuales y pasados de Y, se constata que no participa en ellos ningún otro tipo de variables regresoras y precisamente por eso se dice que " los datos hablan por si mismos" (Gujarati, 2004: 812).

Si después, con fines comparativos se corre el modelo **SES** con MCO y VAR considerando *un número diferente de rezagos* de cada variable, entonces el investigador dispondrá de resultados con MCO y VAR, ¿Cuál usará para predecir? La elección se hará con base en los valores que tengan las estadísticas calculadas, por ejemplo con Eviews. Por consiguiente, en el caso de los criterios de AKAIKE y SCHWARZ escogerá aquel de los criterios que proporcione los valores más bajos de los dos; igualmente, en el caso de las varianzas, seleccionará el modelo que tenga una menor varianza, etc.

En opinión de Pulido et al, (1999, 178): " Estrictamente hablando los modelos VAR típicos no deben incluir variables exógenas, excepto que estas sean de tipo trivial, término constante, tendencia, variables ficticias, etc. No obstante Eviews admite la inclusión de cualquier tipo de variable exógena, estimándose entonces un modelo híbrido VAR-Estructural.

XII.1.2. Bondades, alcance y limitaciones de los modelos VAR

1.- Sus promotores comentan que sus ventajas consisten en que: a).- es un método sencillo de estimación de parámetros, en el que el investigador no tiene porqué preocuparse determinando cuáles son variables endógenas y cuáles son exógenas, puesto que todas son endógenas. Eso sí, él debe saber que en ciertos casos estos modelos VAR incluyen variables puramente exógenas, *en especial, cuando se analiza el efecto de factores estacionales; b).- la* estimación de sus parámetros se hace con todos los términos rezagados de las variables usando MCO en cada ecuación por separado del SES; y c).- En muchos casos las predicciones obtenidas con los modelos VAR son mucho más precisas.

2.- hablando concretamente del software, diremos que una de las utilidades que incorpora Eviews dentro del menú de herramienta del objeto VAR, es la de realizar un Test de Johansen para identificar las posibles relaciones de cointegración entre las distintas variables que incorpora dicho Modelo (Pulido et al (1999: 189)

3.- Sus críticos comentan que tienen los siguientes problemas:

a).- A diferencia de un modelo de ecuaciones simultáneas típico que usa variables endógenas y exógenas, *el modelo VAR es ateórico* porque no explica una teoría económica adecuadamente dado que establece la relación o dependencia de la variable regresada de sí misma rezagada y, que, en consecuencia, al omitir otras variables, no establece relaciones con ellas, es decir, no se construye ninguna teoría, que en su momento, podría expresarse matemáticamente y probarse estadísticamente;

b).- Derivado de esas limitaciones, por su énfasis en la <u>predicción</u> con base univariante, los modelos VAR no son los apropiados para el análisis de políticas públicas o privadas, que generalmente involucran dos o más variables;

c).- El mayor reto en su diseño es determinar la <u>longitud apropiada de los rezagos</u>

de cada variable en cada ecuación del SES. Ejemplo, supóngase que se tiene un modelo VAR de cuatro variables endógenas y se decide que cada una de ellas sea rezagada seis veces en cada ecuación. El resultado es que se tendrán 24 parámetros rezagados en cada ecuación, y si le agregamos el término constante, la ecuación contará con 25 parámetros para su análisis e interpretación correspondiente, situación que hace engorrosa su manipulación económica *sobre todo cuando se alternas los signos de los coeficientes*; además de que estadísticamente ello consumirá demasiados grados de libertad con las consecuencia que ello acarrea.

d).- Si en un modelo VAR la condición es que todas las variables sean estacionarias en *forma conjunta*, cuando no es el caso la información deberá transformarse apropiadamente, digamos diferenciándola. Ello en opinión de Harvey (Gujarati, 2004:827) no siempre produce resultados adecuados; también comenta que cuando el modelo VAR contiene variables I(0) y I(1), variables estacionarias y no estacionarias, la transformación de ellas no es tarea sencilla.;

e).- Los coeficientes individuales estimados no son fáciles de interpretar en virtud de que estos analistas usan frecuentemente la llamada *función de impulso- respuesta (FIR),*que se basa en el estudio de la respuesta de la variable endógena ante shocks en las U_t.

f).- Gujarati en particular (2004:827) termina sugiriendo que el investigador analice también “otras técnicas de predicción” y generaliza señalando que no existe ningún método perfecto que se aplique apropiadamente en todas las situaciones, dado que todos tienen ventajas y desventajas en su aplicación casuística.

g).-Finalmente y derivado de lo anterior con el fin de dar a conocer otro método de predicción que amplía el acervo ya existente, este autor introduce los **modelos ARCH y los GARCH** cuya aplicación sugiere se haga en el estudio de las *series temporales financieras* (precio de las acciones, tasas de interés, de inflación y de cambio, etc.). Una de las características de estos modelos es que la varianza de U_t posiblemente esté correlacionada en el tiempo por la acumulación de la volatilidad que tienen estas variables debido a su naturaleza altamente cambiante en periodos cortos de tiempo. <u>En resumen es pues una opción más para hacer planeación al predecir el valor de una variable en el tiempo.</u>

XII.2.-Los MODELOS VEC: MODELOS DE CORRECIÓN DE ERROR.

Tienen VEC como acrónimo de sus siglas en inglés; MCE, acrónimo de sus siglas en español, significa: Modelo de corrección del error.

<u>Estos modelos son una particularidad de los VAR, son</u> modelos con vectores de cointegración *(Vector Error Correction)* cuyas características básicas abordaremos <u>para propósitos didácticos de la manera siguiente:</u>

En principio diremos que el Instituto L.D. Klein los ubica (UAM, 2004) dentro de las técnicas de predicción, es decir <u>en el ámbito de los modelos VAR</u> cuya sustentación *emana en parte de la metodología de Box & Jenkins que hizo posible la modelización univariante de series de tiempo; mismas que al pronosticarles sus valores futuros suponemos que son estacionarias. Al respecto,* ***ahora*** *es conveniente decir que cuando inicialmente se corre una regresión con MCO de* ***variables*** *no estacionarias se corre el riesgo de obtener correlaciones espurias pero; sin embargo, aquí surge la excepción a la regla anterior, si las series de tiempo no estacionarias son cointegrables, al verificarse que si son cointegradas, enseguida al realizar su regresión, se elimina la correlación espuria entre ellas y, algo adicional muy importante, se obtiene mucha información sobre sus relaciones de equilibrio en el largo plazo. Esto último es muy importante cuando se manejan variables económicas ya que fundamenta su planeación a largo plazo.*

En este contexto es que aparecen los modelos **VEC** como una particularidad de los VAR, **restringidos usualmente a dos variables**; se utilizan para modelar series temporales que son no estacionarias pero que se sabe que son cointegradas, tal que al correr su regresión se obtienen los dos beneficios apuntados en el párrafo anterior.

Derivado de lo anterior, el investigador al correr los modelos VAR sabe que logrará una relación de equilibrio de las variables económicas en **el largo plazo**, pero también está consciente de que en el **corto plazo** pueden existir desequilibrios entre ellas. Aquí surge la conveniencia de utilizar los modelos VEC ya que la proporción del desequilibrio también llamado **error**(dispersión de los términos de las series temporales de la ruta de equilibrio a largo plazo)observado en el corto plazo **se corrige gradualmente mediante ajustes parciales en ese lapso,** con los modelos VEC; esa es la razón por lo que en español se les llame modelos **MCE: Modelos de corrección del error,** dado que reencauzan el alejamiento o variación de sus términos del corto plazo hacia la ruta de equilibrio a largo plazo.

En este sentido también es interesante decir que Pérez (2006:635) con un ejemplo numérico en que usa Eviews deriva el VEC de un VAR, con una interpretación ampliada con respecto a lo que antes explicó el Instituto L.D. Klein tanto en el sentido estadístico como en el económico.

Así, partiendo de un VAR se usan los modelos VEC para conciliar el equilibrio de las variables económicas en el corto con el largo plazo; la condición es que ellas estén cointegradas, situación que se verifica con la Prueba de Johansen y cuya aplicación ilustraremos en los ejercicios siguientes. Mientras tanto conviene profundizar en el tema señalando que Engle y Granger (1987) consideran que una combinación lineal de dos o más series no estacionarias puede ser estacionaria mediante el proceso de diferenciación. La combinación estacionaria puede interpretarse como la cointegración o equilibrio de la relación entre las variables. Se reitera que un modelo VEC es un modelo VAR restringido habitualmente a dos variables**. El planteamiento del VEC restringe la especificación del comportamiento de largo plazo de las variables endógenas a converger a su relación de equilibrio de largo plazo y permiten la dinámica de corto plazo.**

En este sentido conviene citar el ejemplo del Dr. Anyul Insel (2007): Téngase en cuenta la relación entre el consumo y los ingresos en un simple modelo de corrección de errores:

$$\Delta C_t = \theta_1(C_{t-1} – \lambda Y_{t-1}) + U_{1t},\ \theta_1>0\ \ldots\ldots\ldots(1)$$
$$\Delta Y_t = - \theta_2\ (C_{t-2} – \lambda Y_{t-2}) + U_{2t},\ \theta_2>0\ \ldots\ldots(2)$$

Donde u_{1t} y u_{2t} son perturbaciones de ruido blanco, θ_1 , θ_2 representan la velocidad con que se ajustan los valores de las las variables C_t y Y_t las cuales son dos variables endógenas; λ es el corrector en el corto plazo . También, θ_1, θ_2 y λ son parámetros positivos como están denotados en las ecuaciones (1) y (2). El término de cointegración (C_{t-1} - λY_{t-1}) es el término de corrección de errores puesto que la deviación del equilibrio de largo plazo está gradualmente corregido a través de ajustes de corto plazo. Así, como se ha venido reiterando, en un modelo de corrección de errores, la dinámica de corto plazo de las variables en un sistema está influenciada por las desviaciones de equilibrio de largo plazo.

Luego entonces (Insel) en el modelo VEC si:

Las desviaciones son positivas, es decir, (C_{T-1} - λY_{t-1})> 0.

$C_{t-1} = \lambda Y_{t-1}$, cambia C_t y Y_t en respuesta a los choques de u_{1t} y u_{2t}.

Si θ_1 es grande, entonces C_t muestra una mayor respuesta de la desviación de equilibrio de largo plazo que en el periodo anterior.

Cuando θ_1 es pequeño, entonces C_t no responde a las desviaciones de equilibrio del periodo anterior.

Cuando $\theta_2 = 0$, entonces Y_t cambia en respuesta a u_{2t}, por $\Delta Y_t = u_{2t}$. Por lo tanto C_t cambia para eliminar cualquier desviación del equilibrio de largo plazo.

También cuando θ_1=0 ó θ_2 = 0, se dice que no hay relación de causalidad entre las variables cointegradas.

θ_1=0, θ_2 =0, no sería una relación de equilibrio de largo plazo entre las dos variables.

La presentación y descripción de los modelos VAR y VEC constituyen el referente básico del siguiente tema de cointegración, puesto que este concepto también trabaja con dos variables (en su acepción más sencilla, la introducción del tema), <u>cuyos valores de ambas también se determinan mutuamente en un contexto de convergencia a su variación equilibrada, estable (estacionariedad) en horizontes temporales de corto y largo plazo</u>.

XII.3.-COINTEGRACIÓN (Granger).

Como se indicó, los conceptos antes expuestos ahora sirven como referencia para introducir otros más, así como para que con su uso conjunto ahora nos permitan profundizar en el estudio de la relación lineal de dos variables o series temporales, digamos Y_t , X_t .

XII.3.1.- Significado

Al respecto, se habla de cointegración cuando existen dos series que no son estacionarias pero que están relacionadas linealmente, de manera que si les sacamos sus primeras diferencias las hacemos estacionarias y decimos que están cointegradas. Esta estacionariedad le da estabilidad a su variación conjunta en el tiempo, situación que fundamenta la relación de equilibrio de sus variaciones en el corto con el largo plazo. En otras palabras, esta teoría y su metodología proponen la solución al problema de integrar la dinámica a corto plazo con el equilibrio a largo plazo, **de dos variables en estudio. Para ello antes es esencial verificar que las dos series están cointegradas.**

XII.3.2.- Pruebas que se hacen para verificar si existe cointegración de X_t con Y_t,:

a).-Las pruebas de raíces unitarias son para verificar que Y_t , X_t . sean I(1): cointegración de primer orden.

b).-Se aplica la regresión de Y_t sobre X_t y viceversa y se considera $U = y - \hat{\ } x$

c).- Enseguida se aplican las pruebas de raíces unitarias sobre U .

d).- Cuando **ρ** tiende a 1, ello indica que U tiene raíz unitaria y hace suponer que no existen relaciones de cointegración entre Y_t , X_t , por lo que no se identifican relaciones estables entre ellas en el largo plazo.

e).- Como antes se indicó, también se usa la prueba de Johansen

Así si X_t, y Y_t son cointegrables $U = y - \hat{\ } x$ es I(0). Si no lo son, U será I(1). Como las pruebas de raíces unitarias se aplican sobre U, la H_0 es que existe una raíz unitaria, luego en las pruebas de cointegración las H_0 y H_a son:

H_0: U tiene una raíz unitaria ó **ρ** tiende a 1: X_t ,y Y_t no son cointegrables.

H_a: X_t y Y_t son cointegrables ó **ρ** tiende a 0, X_t ,y Y_t son cointegrables.

Como U no es observable se utiliza el residual estimado e al cual aplican las pruebas de ADF para probar la raíz unitaria en U ; *cuando existe más de una raíz unitaria se usa la prueba de Johansen* (Maddala 1996,pág. 673-676).

XII.3.3.- Propiedades

a).Esta metodología se aplica en el **análisis conjunto de dos o más series integradas**, que tienen la propiedad de que su primera *diferencia puede ser estacionaria*. Dicha propiedad se conoce como cointegración (Granger, 2003). Ejemplo , dos series suaves pueden moverse y evolucionar de manera similar aunque no idéntica en torno a un parámetro determinado (μ), en ese caso se dice que son *estacionarias* porque la *distancia* que las separa, es decir su diferencia con respecto a μ es más o menos constante (estable). Luego la cointegración es el procedimiento mediante el cual la diferencia entre dos series integradas (no estacionarias) puede hacerse estacionaria, es decir, moverse con *cambios constantes*, no abruptos alrededor del parámetro μ. Algo semejante se dijo en el Capítulo 10 cuando la estacionariedad se detectaba analizando las fluctuaciones de los valores de los términos de la serie temporal en torno a la varianza.

Otras propiedades útiles e interesantes de las variables son:

b) Deben estar integradas ambas con el mismo factor oculto.

c) Puede considerarse que han sido generadas "*como un modelo de corrección de error"*.

d) Cuando hay predicción con la cointegración digamos cuando sus valores se proyectan hacia el futuro, las predicciones de las dos variables forman un ratio constante porque los parámetros de la ecuación de regresión son lineales y constantes (Granger, 2003). Este resultado asintótico lleva a que esta clase de modelos sea útil porque se usa para determinar equilibrios a largo plazo.

Estas propiedades identificadas por Clive W. Granger y derivadas del análisis conjunto de las variaciones de dos o más series, fue un gran avance dado que Box & Jenkins propusieron métodos para **hacer pronósticos de una única serie temporal integrada a la vez**.

En este sentido es que cuando se habló de una sola serie temporal, se dijo que es ***estacionaria*** *en momento uno (media aritmética) y momento dos (varianza) constantes. Lo anterior se demuestra en el Capítulo de Momentos inserto en mi libro de "Estadística Aplicada al Análisis Económico" (2011).*

Es importante reiterar que para que haya cointegración de dos series integradas (con un comportamiento suave durante su evolución en el tiempo), *éstas deben poseer la propiedad de que una relación lineal de ellas* sea estacionaria.

Al verificar que dos variables tienen la propiedad de la cointegración, se intuye que estas variables tienen las otras propiedades, es decir: a).- que ambas deben estar cointegradas con el mismo factor común oculto; b).- puede considerarse que ambas han sido generadas por el "modelo de corrección de error", **MCE**, en el que las variaciones en una de las series se explican en función de los *retardos de la diferencia entre las dos series,* posiblemente después de un cambio o ajuste de escala, y los *retardos de las diferencias de cada serie*.

Al respecto, es interesante mencionar que dicho modelo fue inventado por el económetra DENNIS SARGAN con el fin de convertir la idea de la cointegración en algo útil en términos prácticos. Para ello seleccionó algunas ecuaciones famosas de la teoría del crecimiento económico y las transformó en estocásticas. Se usa mucho en la modelización macroeconómica para simular políticas públicas, para la evaluación de sus efectos y de sus predicciones, con indicaciones calculadas de la incertidumbre que las rodea. En sí, plantear el análisis usando este modelo sirve para explicar y eliminar diversas dificultades que surgen en el proceso de investigación, como es el caso de las REGRESIONES ESPURIAS.

Granger y Engle (GE) desarrollaron el **MCE** con el fin de conciliar la evolución de corto plazo de una variable económica con su evolución en el largo plazo.

Así, podemos agregar a la metodología descrita párrafos arriba, que cointegración significa que a pesar de que dos o más series temporales a nivel individual sean no estacionarias, su combinación lineal puede ser estacionaria. Para determinar si dos o más series están cointegradas se aplican las pruebas (GE), (AGE), (RCDW) y (MCE), principalmente.

Luego entonces la cointegración de dos o más variables indica que es factible que exista entre ellas una relación de equilibrio en el largo plazo, y que si no existe esa relación de equilibrio en el corto plazo, es decir que haya un desequilibrio en esa relación en el corto plazo, éste, también llamado error, puede eliminarse en el corto plazo aplicando con **MCE.**

Por consiguiente, la cointegración de series temporales hace que aun no siendo ellas estacionarias en sus niveles, al diferenciarse en primer orden, se conviertan en estacionarias.

A manera de resumen de lo anterior puede decirse que la cointegración así entendida y aplicada, es útil en la modelización de **grupos de series**, lo cual significa una superación del enfoque de Box & Jenkins que, como antes se indicó, inicialmente sólo se interesaron preferentemente en el análisis de una única serie integrada a la vez (Granger, 2003). Conviene aclarar que posteriormente la metodología de Box & Jenkins fue generalizada para estudiar dos o más series a la vez mediante el desarrollo de la metodología de los **modelos VAR.**

Así, con la cointegración se formulan modelos matemáticos que relacionan variables económicas que, simultáneamente, ajustan mejor los datos disponibles de las variables. Por su uso, destacan los modelos de planeación que hacen las predicciones a corto y largo plazo sobre el ingreso, el desempleo, el consumo, la inversión, etc., cuyos datos generalmente corresponden a series integradas que deben diferenciarse para hacerlas estacionarias.

En este contexto, otra aportación de Granger es su análisis de la CAUSALIDAD, mismo que le permite afirmar que ésta tiene dos componentes:

1. La causa ocurre antes del efecto, y

2. La causa contiene información sobre el efecto que es único, y no está en otra variable.

El alcance de esta afirmación “es que la variable causal puede contribuir a la predicción de la variable efecto después de que se hayan utilizado previamente otros datos”. Da la impresión de que la causalidad tiene un papel fundamental en la generalización del modelo de corrección del error, sin embargo, es cauteloso al decir que esto pertenece todavía al terreno de la investigación que se está realizando (Granger, 2003).

XII.3.3.1.-Características de dos o más series temporales.

1. Si ambas son estacionarias, su relación se estima con los métodos habituales: MCO o GLM. Al respecto, utilizando el rigor técnico estadístico, se dice que una serie temporal es estacionaria cuando su media aritmética (μ), σ^2 y Covariancia son constantes o invariantes en el tiempo de estudio de la serie temporal.

2. Cuando las series son no estacionarias de orden distinto, no es posible estimar la relación entre ellas.

Las series se pueden estacionarizar (lograr que la serie tenga un comportamiento suave, regular, sin variación significativa estadísticamente a lo largo del tiempo, de manera que sean constantes su media, variancia y covariancia) ya sea aplicando logaritmos, diferencias o ratios con otras variables o hacer una regresión por primeras diferencias, etc.

3. Cuando las series son no estacionarias pero están cointegradas, se puede usar, MCO o GLM, para estimar los efectos a largo plazo y el MCE (modelo de corrección del error) para estimar los efectos a corto plazo.

XII.3.3.2.-El modelo de corrección del error, MCE o VEC en inglés.

4.- Derivado de este último numeral y abundando sobre el MEC, diremos que éste modelo es un modelo VAR restringido que se usa para predecir variables no estacionarias que de antemano sabemos que están cointegradas. *El VEC tiene relaciones de cointegración construidas dentro de la especificación que éste restringe el comportamiento en el largo plazo de las variables endógenas a que converjan a su relación de cointegración,* ***en tanto que permite ajustes dinámicos en el corto plazo*****: (Users Guide EViews 4.0, página 733).**

Cuando se comprueba que están integradas las variables del mismo orden, se elimina la posible existencia de una *relación* espuria entre ellas, relación que además es estacionaria, i.e., que vincula el equilibrio de las variables económicas en el corto y en el largo plazo (Pérez, 2008: 91).

Por lo antes expuesto, debemos decir que el término de cointegración también se conoce como "término de corrección del error" dado que se corrige la desviación o error de corto plazo gradualmente a través de una serie de ajustes parciales en ese lapso, reencauzando así las variaciones ya ajustadas hacia la senda de equilibrio en el largo plazo.

Para ilustrar lo anterior, tomemos el ejemplo de un sistema de dos ecuaciones con una ecuación de cointegración sin términos rezagados ni diferenciados. La ecuación de cointegración es (Pulido et al, 1999: 390):

$$Y_{2t} = \beta Y_{1t}$$

La cual indica que la relación de equilibrio entre las variables sólo se cumplirá a largo plazo. Al respecto, *β es el término de corrección del error.*

Luego el error (desviación) a ir corrigiendo gradualmente es: $Y_{2t} - \beta Y_{1t}$

Su modelo VEC correspondiente es:

$$\Delta Y_{1t} = a_1((Y_{2t-1} - \beta^* Y_{1t-1}) + \varepsilon_{1t}$$

$$\Delta Y_{2t} = a_2((Y_{2t-1} - \beta^* Y_{1t-1}) + \varepsilon_{2t}$$

Como puede observarse en este modelo sencillo pero ilustrativo muestra a β en el lado derecho de la ecuación, que en equilibrio de las dos variables en el largo plazo es igual a cero; sin embargo, si Y_1 e Y_2 se desvían del equilibrio en el largo plazo, β toma valores distintos de cero y cada variable se ajustará parcialmente hasta restablecer la relación de equilibrio entre ellas. **Aquí los coeficientes a_1 y a_2 miden o expresan la velocidad de ajuste de cada variable endógena hacia el equilibrio**.

Comentarios:

1.- En opinión de las personas que se dedican al estudio de las predicciones, es fundamental corregir, eliminar o comprobar estas desviaciones de las variables, ya que limitan de manera significativa la precisión de los pronósticos, situación que en el caso de la planeación es muy importante hacerlo para poder crear escenarios confiables, en corto, mediano y largo plazo;

2.- La relación entre el concepto de cointegración y el MCE para las variables endógenas en estudio, es transcendente porque:

a).- La <u>cointegración da sustento estadístico e interpretación económica al concepto de equilibrio,</u> útil para la formulación del MCE, dado que la conecta con el concepto de equilibrio estadístico como con las variaciones en torno al mismo en el corto plazo;

b).- La modelización del equilibrio del corto con el largo plazo en cierta manera resuelve el debate relativo al uso de variables en niveles o en diferencias, ya que una forma de estacionarizar una serie no estacionaria es “diferenciándola” tantas veces como sea necesario para hacerla estacionaria. Como se recordará, cuando se vio la metodología de Box & Jenkins, este método ha sido criticado porque elimina datos de largo plazo que pueden ser de utilidad para el investigador. Por lo que una opción son los modelos dinámicos, uno de los cuales es el MEC.

c).- Esta vinculación de la cointegración entre las variables endógenas con su orden de integración, reduce o evita la existencia de la relación espuria entre ellas (Luego entonces cuando no integran, decimos que su relación es espuria). Además de que el análisis de cointegración entre series temporales no estacionarias permite distinguir entre las tendencias deterministicas y estocásticas, así como sus respectivas aplicaciones en la inferencia estadística (Pérez, 2008:93)

XII.3.4.- Metodología sugerida para el uso de modelos VAR y VEC

1.- Se establecen el objetivo y el marco teórico de la variable dependiente para poder especificar el modelo y representarlo matemáticamente, así como para probar estadísticamente que sí existe relación entre dicha variable y sus variables explicativas, es decir, para verificar que la ecuación de regresión expresa de manera sencilla la teoría establecida por el investigador.

2.- Al estimarse la ecuación de regresión es conveniente el uso de métodos que determinen el orden de integración de las series y el posible problema de regresión espuria o el sesgo de los estimadores. Dentro de estos métodos destacan el de Granger y Newbold, así como el procedimiento de Johansen que permite identificar la presencia de cointegración, el uso del VAR (modelo de autorregresión vectorial).

3.- Para no generar la regresión espuria (regresión de una serie de tiempo no estacionaria sobre otra estacionaria), ambas series se someten de manera individual a un análisis de raíz unitaria. Si son I (0) son estacionarias y no existe ese riesgo.

4.- Se debe tener cuidado en la interpretación de los resultados, ya que por ejemplo si la evidencia empírica muestra que las variables de la ecuación de regresión tienen una media y una varianza que no son constantes a través del tiempo, ello significa que <u>no son estacionarias</u>, situación que obliga a usar métodos econométricos especiales como la ADF (prueba de raíces unitarias de Dickey y Fuller Aumentado).

XII.3.4.1.-Modelo de Largo Plazo

Cuando las series de variables son estacionarias, el marco natural de análisis es de COINTEGRACIÓN: se específica un modelo VAR sin restricciones de acuerdo con el procedimiento de Johansen. Estos vectores se interpretan como relaciones de equilibrio de largo plazo entre el conjunto de variables bajo estudio. Las principales pruebas estadísticas en el VAR son: La Correlación serial (LM), heterocedasticidad (ARCH), F, la prueba de normalidad Jarque-Bera X y la de White. X, al igual que el análisis de correlación entre los valores actuales y estimados.

XII.3.4.2.- Modelo de corto plazo.

Su construcción es para conocer la dinámica de ajuste de la variable dependiente. Para ello se usa el teorema de Granger (Eagle y Granger) que utiliza el vector de cointegración obtenido anteriormente, como ***mecanismo de corrección de errores*** (VEC) en el procedimiento de lo general a lo específico (Hendry). Al analizar los resultados del modelo se puede conocer si los parámetros de la ecuación son estadísticamente significativos con cierto nivel de α, el uso de variables dicotómicas, etc., es decir se hacen pruebas de diagnóstico: autocorrelación, heterocedasticidad, normalidad, R^2 y error cuadrático medio. Si las pruebas son satisfactorias se puede decir que los datos del modelo permiten predecir adecuadamente la variable dependiente.

XII.3.4.3.- ¿Cuál es la relación entre los modelos de corto y largo plazo?

Es la estabilidad que se logra de las fluctuaciones en el tiempo de las variables que los integran, misma que permite hacer planeación objetiva para escenarios con diferentes horizontes temporales. Su diferencia está en el horizonte temporal de planeación; es de corto plazo si es menos de un año; es de largo plazo, si es mayor a un año.

XII.3.4.4.- Ejercicios con Eviews.

Ejercicio 1

METODOLOGIA DE COINTEGRACION

Como se indicó en la Introducción de este trabajo, el mecanismo de enseñanza-aprendizaje se caracteriza entre otras cosas, por la presentación del marco teórico del tema, en este caso la cointegración. Enseguida, una vez que ya se cuenta con el acervo suficiente del tema, este se ilustra con ejercicios con Eviews, en cuya introducción **se reitera** cuidando de no abrumar o saturar al lector, los conceptos básicos del tema. Con ello se espera que el alumno aprenda, primero, en la exposición teórica y segundo, reafirme sus conocimientos durante la ilustración de la teoría con la metodología apropiada.

En este contexto decimos, como se indicó en el marco teórico de esta tema, que la mayoría de las series económicas temporales son no estacionarias, por lo que suele decirse que pueden contener una raíz unitaria (Pulido et al, 1999: 388), por lo que son integradas de orden 1, indicando que en su comportamiento histórico muestran una pendiente creciente o decreciente, situación que no ayuda a pronosticar su evolución satisfactoria en el tiempo y exige un estudio detallado sobre cómo eliminarla. También vimos que una serie temporal es estacionaria cuando su media y varianza no son afectadas por el tiempo. Cuando si es afectada, se le denomina serie no estacionaria. Al respecto, dijimos en el marco teórico que un ejemplo de esta última es MCA: modelo de caminata aleatoria, que con literales se expresa así: $Yt = Y_{t-1} + \varepsilon_t$ las cuales significan lo siguiente: Y es la serie temporal no estacionaria; Y_{t-1} es ella misma rezagada un periodo y ε es el término aleatorio de error con ruido blanco, lo cual indica que este último no contiene autocorrelaciones y por eso es estacionario. En ese sentido la serie Y conviene decir que tiene un valor constante de predicción, sujeto al tiempo y que su varianza es creciente con el devenir del mismo. También se indicó que el MCA es una serie *estacionaria diferenciada* en virtud de que la primera diferencia de Y es estacionaria y se expresa así:

$$Y_t - Y_{t-1} = (1 – L) = \varepsilon_t$$

En ese sentido decimos que una serie diferenciada estacionaria se denomina integrada de orden d y se expresa con I(d) donde d por consiguiente, representa el orden de integración, entendiendo por éste el número de raíces unitarias que contiene la serie Y_t o, dicho en otra forma: es el número de operaciones de diferenciación que debemos de realizar para que la serie temporal se transforme en estacionaria. Así, decimos para el MCA antes señalado que hay una raíz unitaria, que es una serie integrada de orden 1,I(1). En este contexto también dijimos que una serie estacionaria será de orden 0, I(0). Luego una serie temporal Y_t no estacionaria es integrada de orden d y se indica como Y_t tiende a I(d) podemos convertirla en estacionaria diferenciándola d veces. Por su importancia y uso sucesivo que tendrán, es conveniente decir que Granger y Engle (1987) tipificaron o caracterizaron las series I(0) y I(1) de la siguiente manera:

Concepto	Serie I(0)	Serie I(1)
Media	Sus valores fluctúan en torno a la media.	Sus valores oscilan significativamente y no alrededor de la media.
Varianza	Esta estadística es finita e independiente del tiempo.	Esta estadística varía con el tiempo y tiende a infinito a medida que éste transcurre.
Autocorrelación	La autocorrelación entre los términos de la serie disminuye a medida que se incrementan los rezagos.	La autocorrelación tiende a 1 para cualquier orden de rezago.
Efectos de innovación	Tienen memoria limitada, por consiguiente éstos no tienen efectos permanentes.	Cualquier innovación afecta permanentemente a sus procesos.

En el ejercicio anterior también indicamos que las pruebas para detectar si la serie es estacionaria, son los informales (gráfica y correlograma) y los formales (DF, ADF y PP).

Cointegración.

En este contexto es que Granger y Engle (1987) comentan que una combinación lineal de dos o más series no estacionarias, puede ser estacionaria. Luego si existe una combinación lineal de dos o más series que es estacionaria, i.e., I(0), se dice que las series no estacionarias, con una raíz unitaria, mismas que dan lugar a esa combinación lineal, están cointegradas.

Así, la combinación lineal estacionaria se conoce como *ecuación de cointegracion* y bien puede decirse que *es la relación de equilibrio a largo plazo entre las variables que conforman la ecuación.* Un ejemplo clásico de cointegración entre variables lo proporciona la relación que existe entre el consumo y el ingreso de las personas. Si no estuvieran cointegradas (relacionadas a largo plazo), se daría la impresión de que a largo plazo el consumo podría variar permanentemente “sin alcanzar un valor estable de equilibrio” (Pulido et al, 1999: 389).

Por otra parte pero en conexión con el tema, es conveniente reiterar que *un modelo de corrección de error (MCE) o vector de corrección del error (VCE)* es un modelo (VAR) restringido por lo general, a dos variables, que se utiliza para predecir series no estacionarias pero que se sabe (como en el caso anterior) que son cointegradas. Señala Pulido et al que la especificación de un VEC *restringe* la evolución de las variables endógenas en el largo plazo, con objeto de que converjan a sus relaciones de cointegración, en tanto que *admite* “un amplio espectro de dinamicidades a corto plazo”. Luego entonces el concepto de cointegración en ese sentido, debe interpretarse como una corrección del error, entendiendo por error la *desviación* de los valores de las variables respecto al equilibrio a largo plazo, misma que se corrige paulatinamente por medio de ajustes parciales en el corto plazo. *Ejemplo algebraico* de la creación de una ecuación de cointegración considerando dos variables: Y_{1t}, Y_{2t}, sin término de diferenciación alguno y sin incluir valores retardados de ellas en el lado derecho de la ecuación, aun cuando por lo general siempre se incluyen las variables endógenas retardadas: Y_{1t-1}, Y_{2t-1}.

Así, la ecuación de cointegración (Pulido et al, 1999: 390) es: $Y_{2t} = \beta Y_{1t}$. La cual indica que la relación de equilibrio entre las variables sólo se cumplirá a largo plazo. Al respecto, *β es el término de corrección del error.*

Luego el error (desviación) a ir corrigiendo gradualmente es: $Y_{2t} - \beta Y_{1t}$ por lo que el vector de corrección del error (VCE) será:

$\Delta Y_{1t} = a_1((Y_{2t-1} - \beta^* Y_{1t-1}) + \varepsilon_{1t}$, en la que Δ en las dos ecuaciones significa variación; ahora digamos por consiguiente que la segunda ecuación es:

$$\Delta Y_{2t} = a_2((Y_{2t-1} - \beta^* Y_{1t-1}) + \varepsilon_{2t}$$

Obsérvese que: a).- β, el corrector del error aparece en ambas ecuaciones y b).- que β está al lado derecho de la ecuación. En el equilibrio a largo plazo β toma el valor de cero, pero si Y_1 y Y_2 se desvían de ese equilibrio, $\beta \neq 0$ y cada una de esas variables se ajusta parcialmente para restablecer la relación de equilibrio. Ahora diremos que a_1 y a_2 representan la velocidad con que se ajusta gradualmente en el corto plazo la desviación incurrida en el largo plazo.

Observaciones importantes sobre el diseño de la ecuación de cointegración:

1.- Para la construcción de la ecuación hemos *supuesto* que las dos variables endógenas Y_1 y Y_2 tienen una media distinta de cero y que la ecuación de integración no tiene un término independiente (por eso no aparece en la ecuación).

2.- Sin embargo, por otra parte conviene decir que también puede *suponerse* que estas dos variables endógenas no tienen tendencia pero que, ahora si, la ecuación de cointegración tiene un término independiente (μ). En este modelo, el vector de corrección del error, VEC, se representa así:

$\Delta Y_{1t} = a_1(Y_{2t-1} - \mu - \beta^* Y_{1t-1}) + \varepsilon_{1t}$ en la que Δ representa variación ;

$$\Delta Y_{2t} = a_2(Y_{2t-1} - \mu - \beta^* Y_{1t-1}) + \varepsilon_{2t}$$

3.- Otro *supuesto* interesante para la construcción de la ecuación de cointegración en un modelo VEC puede ser: que considere que las dos variables endógenas presenten tendencia y un término constante (δ), luego el VEC se expresará así:

$$\Delta Y_{1t} = \delta_1 + a_1(Y_{2t-1} - \mu - \beta^* Y_{1t-1}) + \varepsilon_{1t}$$ en la que Δ representa variación ;

$$\Delta Y_{2t} = \delta_2 + a_2(Y_{2t-1} - \mu - \beta^* Y_{1t-1}) + \varepsilon_{2t}$$

4.- Puede haber otro tipo de supuestos a considerar para construir la ecuación de cointegración, mismos que por consiguiente, cambiarán las ecuaciones del VEC.

5.- Eviews está diseñado para calcular la ecuación de cointegración con varios de los supuestos anteriores.

Metodología para determinar si las series están cointegradas:

Señala Pulido et al (1999) que para determinar si un grupo de series no estacionarias están cointegradas y, por consiguiente, si es el caso, determinar las ecuaciones de cointegración, que expresan el equilibrio a largo plazo entre las variables en estudio, se puede recurrir a Eviews , mismo que contiene la prueba para verificar si hay cointegración sobre modelos VAR utilizando la metodología que formuló Johansen (1991) y que se caracteriza por analizar las restricciones impuestas por la cointegración de las series incluidas en un modelo VAR no restringido.

Al respecto, cabe señalar que Johansen estableció el siguiente planteamiento teórico considerando un modelo VAR de orden *p*:

$$Y_t = A_t Y_{t-1} + \ldots\ldots\ldots + A_p Y_{p-1} + BX_t + \varepsilon_t$$

Especificando que Y_t es un vector de *k* variables no estacionarias de orden 1, I(1), que X_t es un vector de *d* variables deterministas y que ε_t es un vector de "innovaciones" o término de error.

Así, podemos decir que la metodología de cointegración de Johansen nos permite verificar que las variables endógenas están cointegradas, así como el número de ecuaciones de cointegración. **Dicha metodología se expone a continuación aplicándola al análisis de la relación que existe entre las exportaciones y las importaciones mexicanas:**

Para crear la base de datos (workfile) usando el procedimiento acostumbrado: File/new/workfile/start y end date/ok/ quick/empty group(edit series)/ vaciar allí los datos de las siguientes series, X,M en millones de dólares, correspondientes a la economía mexicana.

Cuadro 12.3.4.4.1. Creación de base de datos en Eviews.

X: Exportaciones	M: Importaciones
5811	7146
6060	8614
6263	9442
6726	10092
8156	10741
8132	12010
6989	12067
7834	12534
7018	11112
7709	10232
8038	8413
8207	7105
7547	5908
8116	6707
8270	7236
8996	7218
9769	7595
9502	8258
9288	8997
9271	8797
9245	9071
8379	8901
9110	8796
9124	8291
7517	7917
7342	8194
7025	7659
8043	7531
8720	7342
9549	8026
9403	8872
9697	8890
10339	9580
10749	10713
10343	11797
10664	12382
11697	12727
12347	13618
11708	13770
12351	13810
13122	15313
12469	14128
14204	15982
16276	18100

13391	15534
14807	18659
14606	18764
15283	19777
14460	19612
15467	21365
15484	22339
16257	22792
15628	21289
16952	22597
16683	23349
18489	23916
18062	24843
19406	26882
19461	27369
21443	28939
23017	24372
24056	23700
24413	24864
25543	25671
26626	26790
28417	28162
29104	29969
31169	32904
30415	30635
32659	33843
33334	35924
34910	38581
33704	36983
35476	38868
34207	38921
36762	41369
35424	39062
39268	42134
40821	43969
43397	47694
44326	49256
47463	50949
49489	53512
51598	57843
46942	51718
47806	51382
45788	49111
45630	51653
43356	47115
48427	51287
48090	51317

48267	52558
46400	48711
47911	49708
49427	51270
51653	54274
51683	53170
57264	56798
57071	58313
58886	63218
56750	59310
64774	64869
64802	65271
70192	72275
69935	69407
75941	75804
74990	75685
76070	78271
72625	74984
79966	81477
82752	83751
86569	87513
82609	84099
92093	94110

Paso 1.- Pruebas de no estacionariedad:

Con los datos anteriores de las dos series temporales, iniciemos el proceso, cuyo primer paso es detectar la estacionariedad de ellas.

Para verificar si X y M son no estacionarias se utilizarán los siguientes métodos:

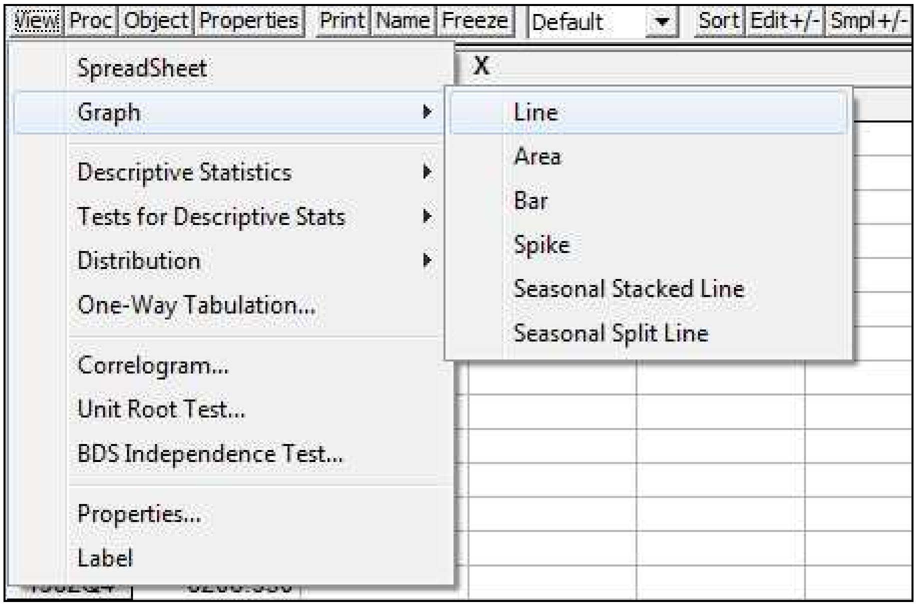
1.- Informales (gráfico y correlograma) y

2.- Formales pruebas Dickey –Fuller , Dickey – Fuller Aumentada y Phillips-Perron

1.- Informales:

Para **X**: En la base de datos (workfile) seleccionamos la variable exportaciones, representada por "X", con click izquierdo aparece un menú, de ese menú se selecciona se selecciona la opción view, posteriormente graph y por último line tal y como se muestra en el siguiente cuadro.

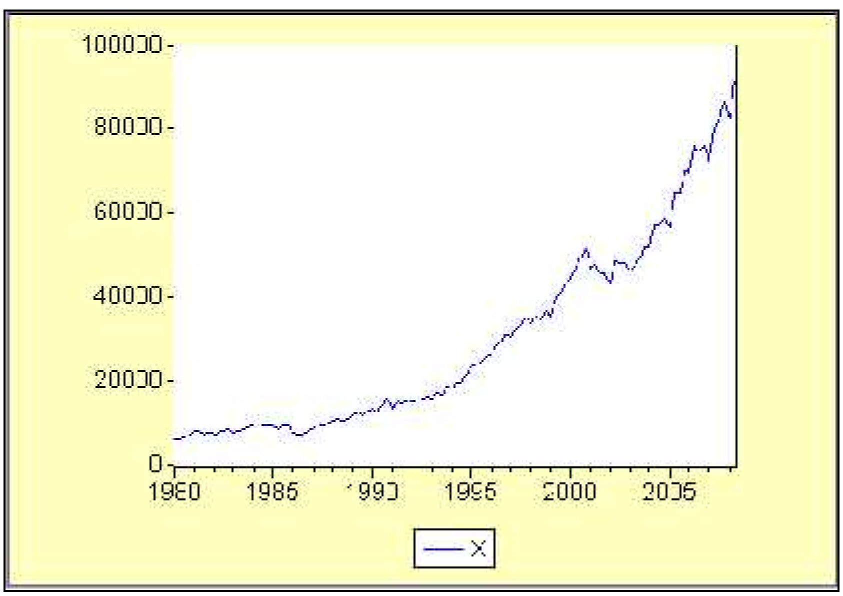
Cuadro 12.3.4.4.2 Procedimiento para obtención de gráfica.



A continuación aparece la gráfica para las exportaciones (X).

a) Gráfica para X:

Gráfica 12.3.4.4.1.: Comportamiento gráfico de las exportaciones.

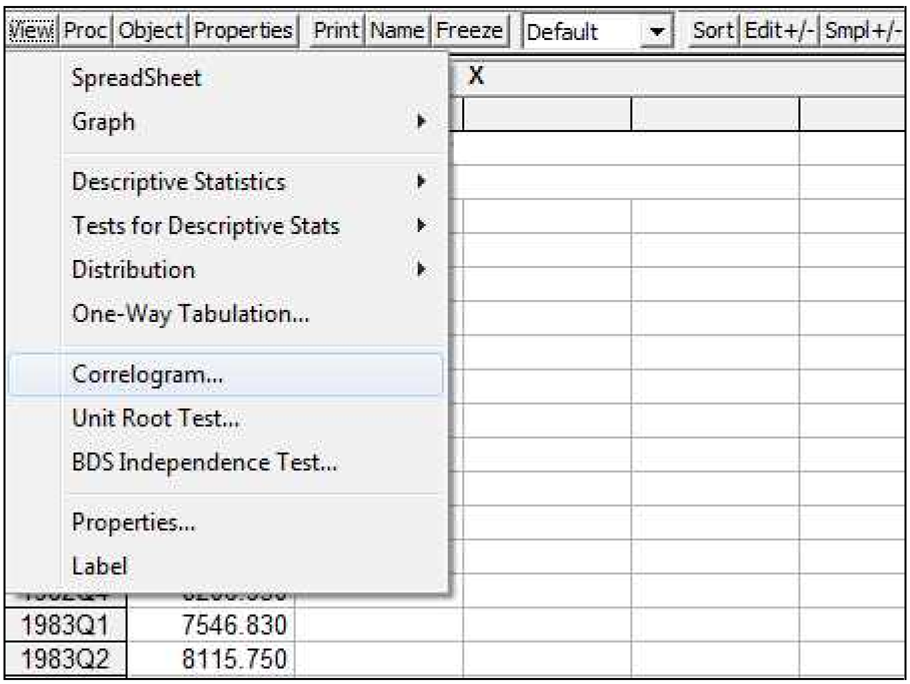


Como puede observarse, existe una tendencia creciente, lo cual revela que conforme la serie evoluciona en tiempos sus términos se alejan de la media aritmética de los mismos y por consiguiente no tienen varianza ni covarianza constante. Es posible que sus valores deambulen en el tiempo como una caminata aleatoria, de ahí que no garanticen constancia en su comportamiento histórico y qué no constituyan una base sólida para proyectarlos al futuro.

Lo anterior no debe sorprender mucho porque se ha observado que por lo general la series económicas son no estacionarias, por lo que de no corregirse, al aplicarse por ejemplo, las teorías convencionales de MCO a series temporales “sin limpiarse” tienden a producir resultados espurios i.e., relaciones que no son reales.

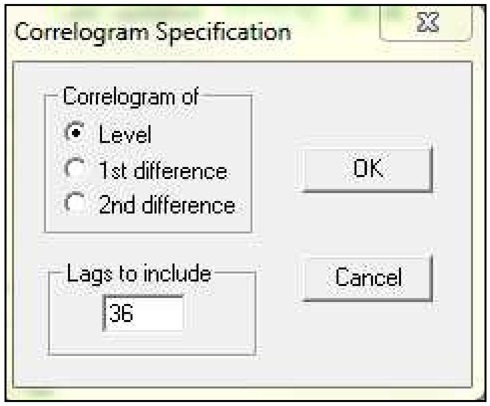
Nuevamente se selecciona de la base de datos (workfile) la variable X, a continuación open, en seguida view y del submenú que se despliega seleccionar la opción correlogram. Se ilustra a continuación.

Cuadro 12.3.4.4.3 Procedimiento para la obtención de correlograma.



Una vez realizada la operación anterior, se especifican las cualidades que contendrá el correlograma, en este caso se seleccionan los niveles y los rezagos a incluir son 36.

Cuadro 12.3.4.4.4 Especificación de correlograma.



Una vez obtenido lo anterior, se obtiene la siguiente tabla.

b) Tabla de correlograma:

Tabla 12.3.4.4.1: Correlograma de las exportaciones.

Correlogram of X

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.963	0.963	107.40	0.000
		2	0.926	0.121	209.02	0.000
		3	0.889	-0.093	303.13	0.000
		4	0.856	0.027	391.33	0.000
		5	0.821	-0.020	470.14	0.000
		6	0.796	0.087	550.67	0.000
		7	0.765	-0.073	622.55	0.000
		8	0.732	0.022	689.45	0.000
		9	0.698	-0.072	750.36	0.000
		10	0.667	0.054	806.83	0.000
		11	0.638	-0.017	858.59	0.000
		12	0.610	0.037	906.82	0.000
		13	0.584	0.045	950.93	0.000
		14	0.561	0.074	992.53	0.000
		15	0.538	-0.013	1030.3	0.000
		16	0.515	-0.015	1066.7	0.000

Conforme el marco teórico lo señala, las hipótesis se plantean así:

Ho: Y_t es estacionaria si la probabilidad asociada a Q es mayor que α=5%;

Ha: Y_t es no estacionaria si la probabilidad asociada a Q es menor que α=5%;

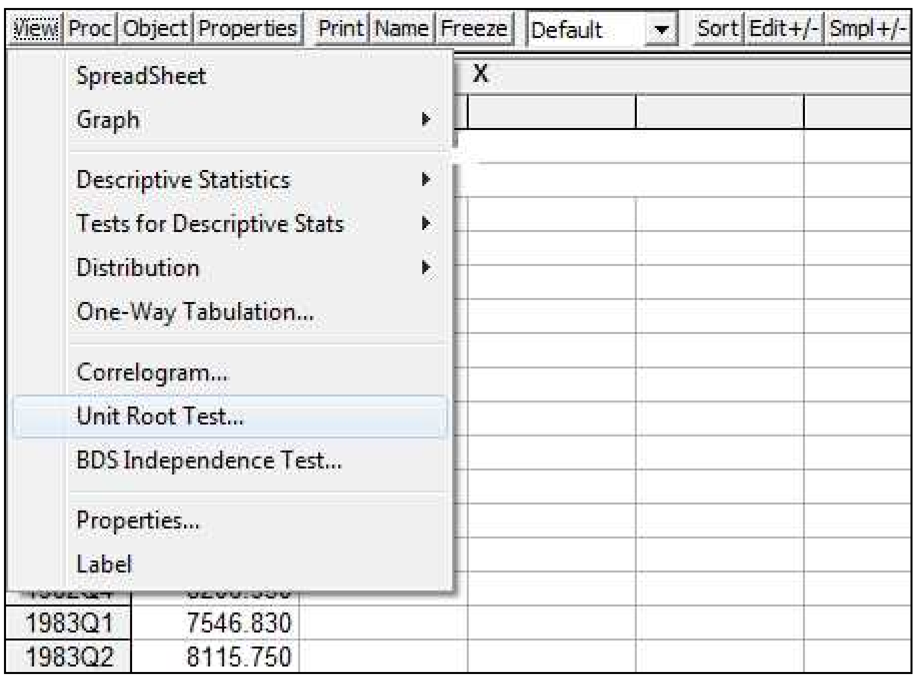
Verificación de hipótesis: AC y PAC *gráficamente* revelan que varias p salen de las bandas de confianza y que estas toman valores cercanos a 1 indicando la no estacionariedad de la serie. AC *numéricamente* no muestra un descenso rápido de las p; PAC muestra que de las ρ´s, sólo la primera es significativamente diferente de cero, indicando que en los primeros rezagos existe alta autocorrelación dado que sus valores están próximos a 1 propiciando con ello perturbaciones que aparecerían en caso de que sus valores fueran proyectados. Las Q tienen una probabilidad de cero indicando la no estacionariedad. Luego entonces grafica y numéricamente hay congruencia y por ello es que rechazamos Ho y aceptamos Ha: Y_t es no estacionaria porque la probabilidad asociada a Q es menor que α=5%;

2.- Pruebas formales:

a).- ADF

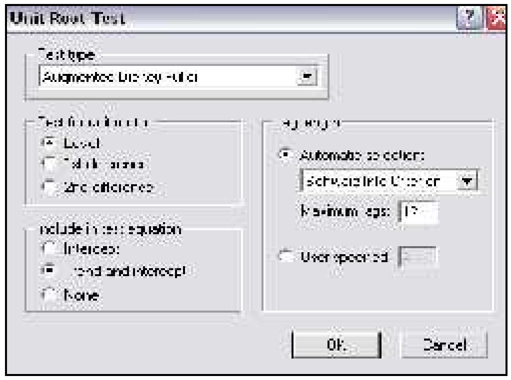
Con el fin de realizar dicha prueba se retoma la serie "X" desde el workfile o base de datos, a continuación se abre la serie, en seguida se abre la opción view y del submenú se selecciona unit root test. Los pasos anteriores se ilustran en el siguiente cuadro.

Cuadro 12.3.4.4.5 Procedimiento para la obtención de prueba ADF.



Y aparece un cuadro cuyo título es "Unit Root Test" que nos pregunta sobre el test a realizar, le decimos que "Augmented Dickey-Fuller/ en él seleccionamos " level" y " Trend & Intercept", dejando igual los datos que Eviews por default incluye/ ok y aparece el siguiente cuadro:

Cuadro 12.3.4.4.6: Raíz Unitaria Prueba Dickey-Fuller Aumentada.



Con ello se obtiene el siguiente cuadro:

Cuadro 12.3.4.4.7: Análisis de regresión de la prueba Dickey-Fuller Aumentada.

Augmented Dickey-Fuller Unit Root Test on X			
Null Hypothesis: X has a unit root			
Exogenous: Constant, Linear Trend			
Lag Length: 12 (Automatic based on SIC, MAXLAG=12)			
		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.512757	0.9992
Test critical values:	1% level	4.051460	
	5% level	-3.454319	
	10% level	-3.153171	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(X)
Method: Least Squares
Date: 03/17/09 Time: 10:43
Sample (adjusted): 1983Q2 2008Q2
Included observations: 101 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	0.012962	0.025279	0.512757	0.6091
D(X(-1))	-0.001375	0.109652	-0.012539	0.9900
D(X(-2))	0.081081	0.106383	0.762157	0.4481
D(X(-3))	0.067522	0.103363	0.653249	0.5153
D(X(-4))	0.238549	0.106625	2.237270	0.0279
D(X(-5))	0.288973	0.109180	2.646792	0.0097
D(X(-6))	-0.109584	0.113567	-0.964928	0.3373
D(X(-7))	0.117464	0.116472	1.008513	0.3160
D(X(-8))	0.234090	0.113170	2.068520	0.0416
D(X(-9))	-0.077001	0.114525	-0.675669	0.5011
D(X(-10))	-0.255039	0.115941	-2.199727	0.0305
D(X(-11))	-0.363115	0.119051	-3.050067	0.0030
D(X(-12))	0.381697	0.125503	3.041327	0.0031
C	256.1793	434.7500	0.589257	0.5572
@TREND(1980Q1)	11.51459	15.68458	0.734135	0.4649

R-squared	0.671141	Mean dependent var	837.7891
Adjusted R-squared	0.617605	S.D. dependent var	2209.784
S.E. of regression	1366.407	Akaike info criterion	17.41410
Sum squared resid	1.61E+08	Schwarz criterion	17.80251
Log likelihood	-864.4105	F-statistic	12.30644
Durbin-Watson stat	1.935153	Prob(F-statistic)	0.000000

Planteamiento de las hipótesis:

Ho. Y_t tiene raíz unitaria, luego es no estacionaria; su probabilidad debe ser mayor que α=5%; Además, el valor ADF que corresponde a la estadística t de Student para X debe ser **menor** en valor absoluto en todos los niveles de significación (1%,5%, 10%) que los sugeridos por McKinnon.

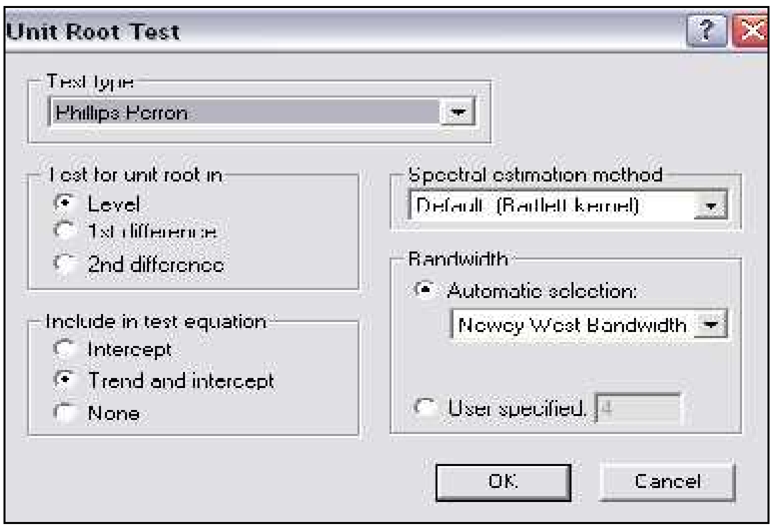
Ha: Y_t no tiene raíz unitaria, luego es estacionaria; su probabilidad debe ser menor que α=5%; Además, el valor ADF que corresponde a la estadística t de Student para X debe ser **mayor** en valor absoluto en todos los niveles de significación (1%,5%, 10%) que los sugeridos por McKinnon.

Verificación de las hipótesis: Al ser el valor t= 0.512757 de ADF **menor** en valores absolutos en todos los niveles de significación (1, 5 y 10%) sugeridos por MacKinnon, ello indica que: aceptamos Ho, es decir, la serie X es no estacionaria, por tanto, tiene raíz unitaria. Asimismo, la probabilidad asociada a t de ADF es mayor a 0.05 lo que indica la no estacionariedad de X.

b).- Prueba PP: Phillips- Perron

Seguimos las mismas indicaciones que en el caso anterior, pero ahora cuando el cuadro nos pregunta sobre el tipo de test a realizar le decimos que el de Phillips-Perron, dejando igual el resto de datos que por default aparecen:

Cuadro 12.3.4.4.8: Raíz Unitaria Prueba Phillips-Perron.



Enseguida aparece el siguiente cuadro:

Cuadro 12.3.4.4.9: Análisis de Regresión de la prueba Prueba Phillips-Perron.

Phillips-Perron Unit Root Test on X

Null Hypothesis: X has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 9 (Newey-West using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		1.122328	0.9999
Test critical values:	1% level	-4.041280	
	5% level	-3.450073	
	10% level	-3.150336	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	4052656.
HAC corrected variance (Bartlett kernel)	2102517.

Phillips-Perron Test Equation
Dependent Variable: D(X)
Method: Least Squares
Date: 03/17/09 Time: 10:46
Sample (adjusted): 1980Q2 2008Q2
Included observations: 113 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
X(-1)	0.003710	0.024488	0.151500	0.8799
C	-264.0983	429.0840	-0.615493	0.5395
@TREND(1980Q1)	16.15062	16.73860	0.964873	0.3367

R-squared	0.082825	Mean dependent var	763.5575
Adjusted R-squared	0.066149	S.D. dependent var	2111.416
S.E. of regression	2040.388	Akaike info criterion	18.10586
Sum squared resid	4.58E+08	Schwarz criterion	18.17827
Log likelihood	-1019.981	F-statistic	4.966721
Durbin-Watson stat	2.726972	Prob(F-statistic)	0.008608

Interpretación: decimos que aceptamos la Ho, que la serie tiene raíz unitaria y por ello X es no estacionaria. El resto de las estadísticas utilizadas para decidir la estacionariedad de la serie muestran relaciones similares a las descritas con ADF, de ahí que concluyamos que la serie es no estacionaria.

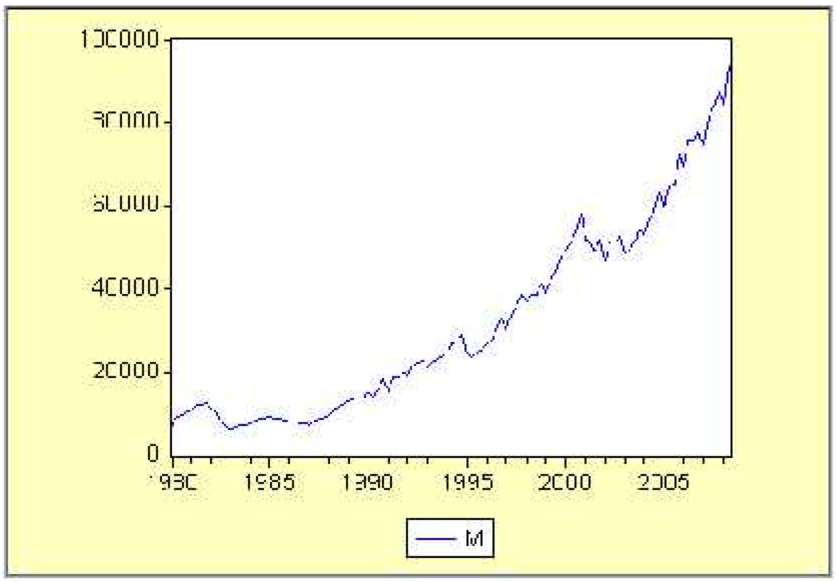
Verificación de la no estacionariedad de M

En este contexto, ahora hagamos la prueba de estacionariedad para la variable M, siguiendo los mismos pasos de X, ahora aparece su

1.- Pruebas informales:

a) Gráfica:

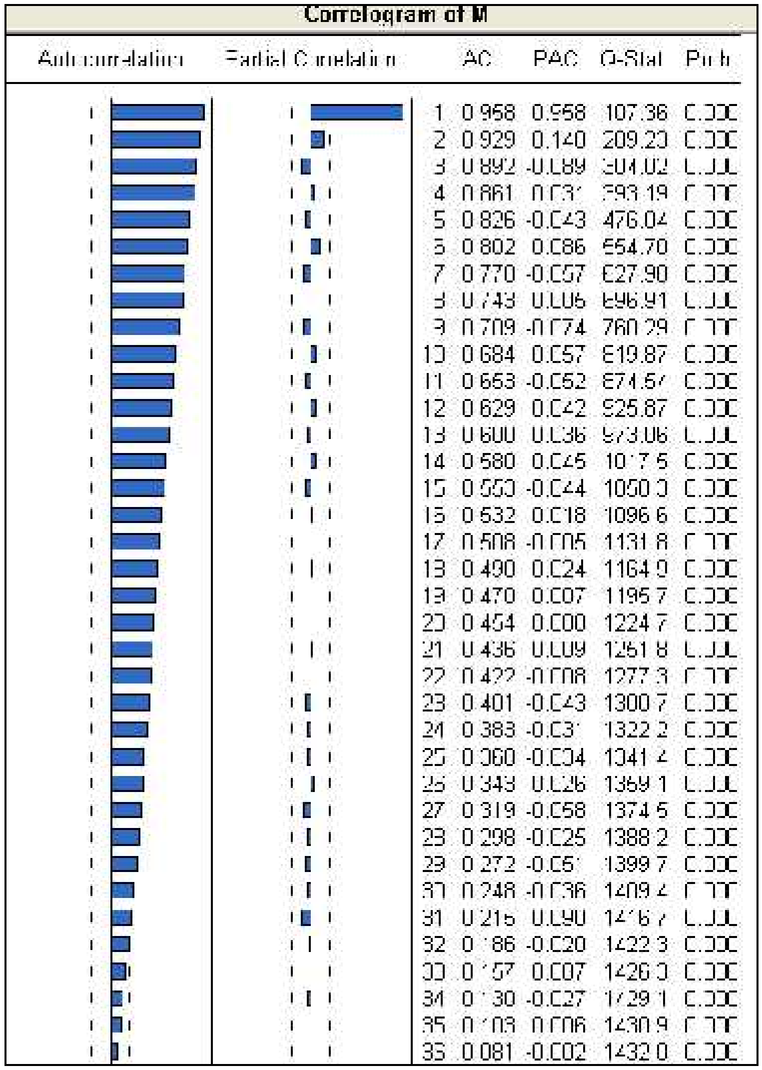
Gráfica 12.3.4.4.2.: Comportamiento de las importaciones.



Claramente se observa que la serie es no estacionaria por la tendencia creciente que presenta. Ello se sustenta en los mismos argumentos estadísticos que describimos antes para la variable X.

b) Correlograma:

Tabla 12.3.4.4.2: Correlograma de las Importaciones.



Correlogram of M

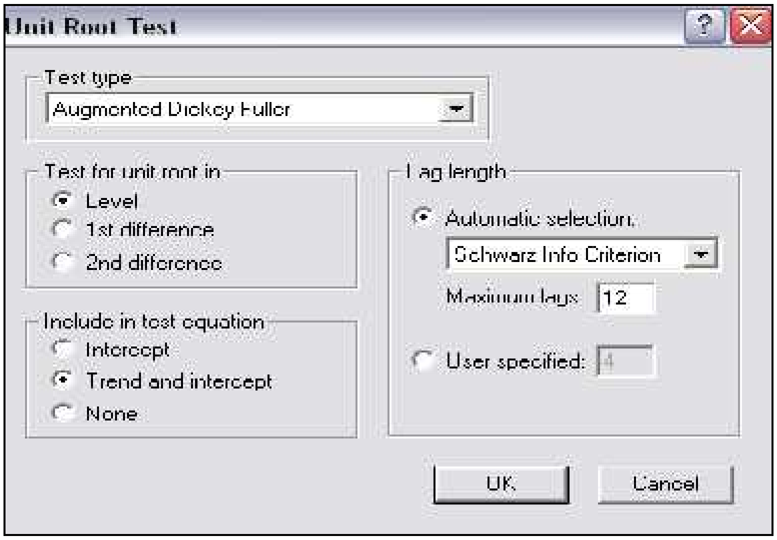
	AC	PAC	Q-Stat	Prob
1	0.958	0.958	107.36	0.000
2	0.929	0.140	209.20	0.000
3	0.892	-0.089	304.02	0.000
4	0.861	0.031	393.19	0.000
5	0.826	-0.043	476.04	0.000
6	0.802	0.086	554.70	0.000
7	0.770	-0.057	627.90	0.000
8	0.743	0.006	696.91	0.000
9	0.709	-0.074	760.29	0.000
10	0.684	0.057	819.87	0.000
11	0.653	-0.052	874.54	0.000
12	0.629	0.042	925.87	0.000
13	0.600	0.036	973.06	0.000
14	0.580	0.045	1017.5	0.000
15	0.550	-0.044	1050.0	0.000
16	0.532	0.018	1096.6	0.000
17	0.508	-0.005	1131.8	0.000
18	0.490	0.024	1164.9	0.000
19	0.470	0.007	1195.7	0.000
20	0.454	0.000	1224.7	0.000
21	0.436	0.009	1251.8	0.000
22	0.422	-0.008	1277.3	0.000
23	0.401	-0.043	1300.7	0.000
24	0.383	-0.031	1322.2	0.000
25	0.360	-0.034	1341.4	0.000
26	0.343	0.026	1359.1	0.000
27	0.319	-0.058	1374.5	0.000
28	0.298	-0.025	1388.2	0.000
29	0.272	-0.051	1399.7	0.000
30	0.248	-0.036	1409.4	0.000
31	0.215	0.090	1416.7	0.000
32	0.186	-0.020	1422.3	0.000
33	0.157	0.007	1426.0	0.000
34	0.130	-0.027	1429.1	0.000
35	0.103	0.006	1430.9	0.000
36	0.081	-0.002	1432.0	0.000

Interpretación similar a la serie X: M es no estacionaria.

2.- Pruebas formales

a).-ADF:

Cuadro 12.3.4.4.10: Raíz Unitaria Prueba Dickey-Fuller Aumentada.



De lo cual se deriva el siguiente cuadro de resultados:

Cuadro 12.3.4.4.11: Análisis de regresión de la prueba Dickey-Fuller Aumentada.

Augmented Dickey-Fuller Unit Root Test on M

Null Hypothesis: M has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 8 (Automatic based on SIC, MAXLAG=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-0.968290	0.9432
Test critical values:	1% level	-4.047795	
	5% level	-3.453179	
	10% level	-3.152153	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(M)
Method: Least Squares
Date: 03/17/09 Time: 10:52
Sample (adjusted): 1982Q2 2008Q2
Included observations: 105 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M(-1)	-0.027281	0.028174	-0.968290	0.3354
D(M(-1))	-0.113997	0.098993	-1.151562	0.2524
D(M(-2))	0.190344	0.100385	1.896152	0.0610
D(M(-3))	-0.019805	0.098006	-0.202084	0.8403
D(M(-4))	0.291141	0.097208	2.995042	0.0035
D(M(-5))	-0.175508	0.100350	-1.748962	0.0836
D(M(-6))	-0.276340	0.102282	-2.701736	0.0082
D(M(-7))	-0.051203	0.104276	-0.491033	0.6245
D(M(-8))	0.452340	0.103945	4.351706	0.0000
C	-693.3589	459.6716	-1.508379	0.1348
@TREND(1980Q1)	35.39603	19.34417	1.829804	0.0704

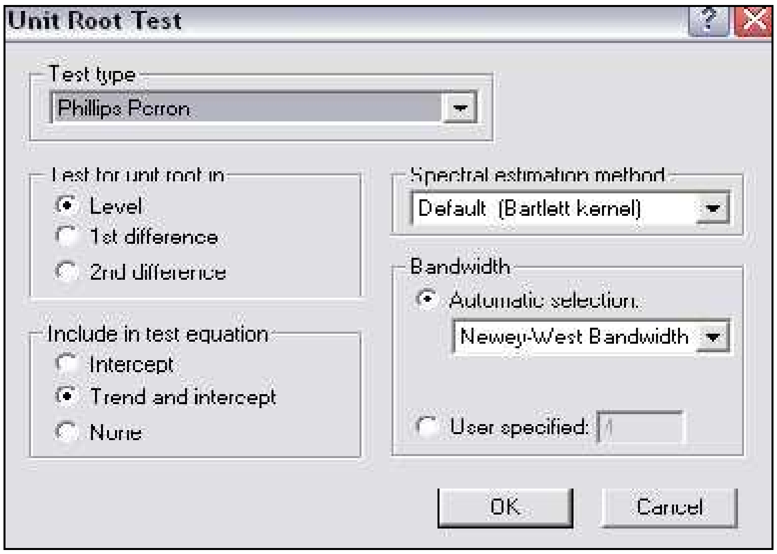
R-squared	0.600154	Mean dependent var	790.4571
Adjusted R-squared	0.557617	S.D. dependent var	2457.161
S.E. of regression	1634.304	Akaike info criterion	17.73468
Sum squared resid	2.51E+08	Schwarz criterion	18.01271
Log likelihood	-920.0707	F-statistic	14.10904
Durbin-Watson stat	1.839076	Prob(F-statistic)	0.000000

El valor de t de ADF en valores absolutos es **menor** a los valores críticos al 1, 5 y 10% de MacKinnon y su probabilidad es **mayor** a 0.05 lo que indica que la serie M es no estacionaria, que aceptamos Ho: M tiene raíz unitaria y es no estacionaria.

b).- Prueba PP

Siguiendo las mismas instrucciones de X, ahora seleccionamos:

Cuadro 12.3.4.4.12: Prueba de Raíz Unitaria Phillips-Perron.



Tras lo cual se obtiene:

Cuadro 12.3.4.4.4.4.13: Análisis de Regresión de la prueba Prueba Phillips-Perron.

Phillips-Perron Unit Root Test on M

Null Hypothesis: M has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 9 (Newey-West using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		0.350975	0.9987
Test critical values:	1% level	-4.041280	
	5% level	-3.450073	
	10% level	-3.150336	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	5259207.
HAC corrected variance (Bartlett kernel)	3191545.

Phillips-Perron Test Equation
Dependent Variable: D(M)
Method: Least Squares
Date: 03/17/09 Time: 10:54
Sample (adjusted): 1980Q2 2008Q2
Included observations: 113 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M(-1)	-0.011164	0.029654	-0.376488	0.7073
C	-328.0088	477.8986	-0.686356	0.4939
@TREND(1980Q1)	25.34006	20.54632	1.233314	0.2201

R-squared	0.062829	Mean dependent var	769.5929
Adjusted R-squared	0.045789	S.D. dependent var	2379.473
S.E. of regression	2324.358	Akaike info criterion	18.36647
Sum squared resid	5.94E+08	Schwarz criterion	18.43887
Log likelihood	-1034.705	F-statistic	3.687246
Durbin-Watson stat	2.695071	Prob(F-statistic)	0.028187

Interpretación similar a la t de ADF, por tanto la serie M es no estacionria: tiene raíz unitaria.

Conclusión: X y M son series no estacionarias, que es la primera condición para cointegrarlas en una ecuación que exprese su equilibrio en el largo plazo.

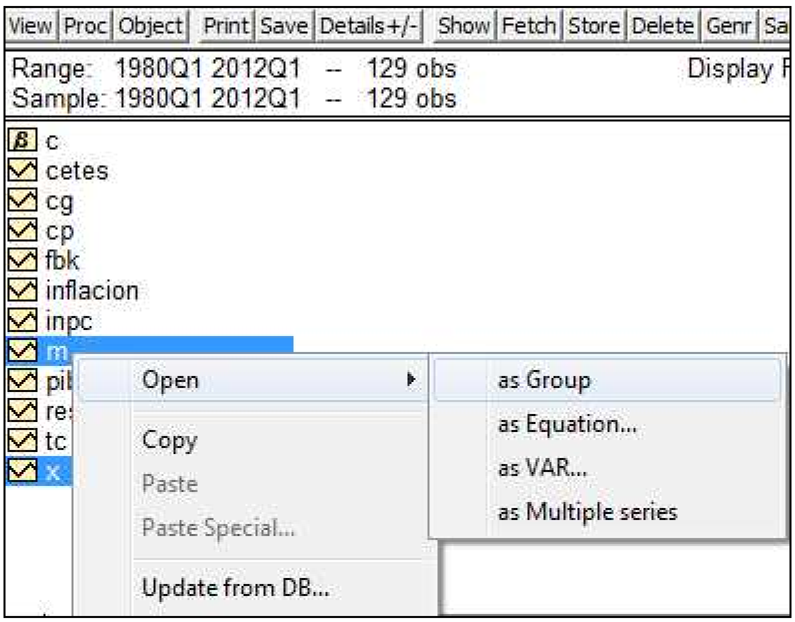
Paso 2.- Relación de X y M en el corto plazo

Con base en lo anterior ahora podemos pasar a establecer la relación en el corto plazo entre las variables. Para ello antes es conveniente averiguar con la prueba de causalidad de Granger si existe o no causalidad entre las dos variables.

Paso 2. 1: Prueba de causalidad de Granger

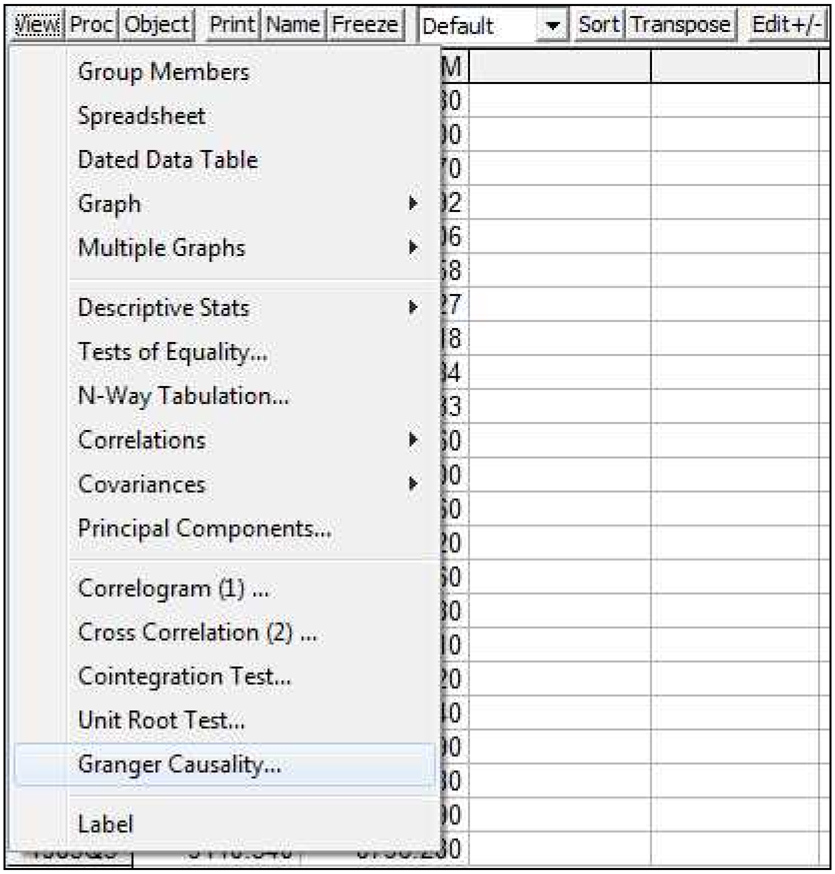
Lo anterior, en cierta forma, se inicia corroborando con la **Prueba de causalidad de Granger**. Para obtenerla en Eviews en el workfile seleccionando las variables endógenas, en este caso X y M, y las abrimos como grupo así: se selecciona X, presionando ctrl se selecciona M, con botón izquierdo se da un click en open y del submenú se elige as group. Se ilustran los pasos a continuación.

Cuadro 12.3.4.4.14: Procedimiento para abrir la serie como un grupo.



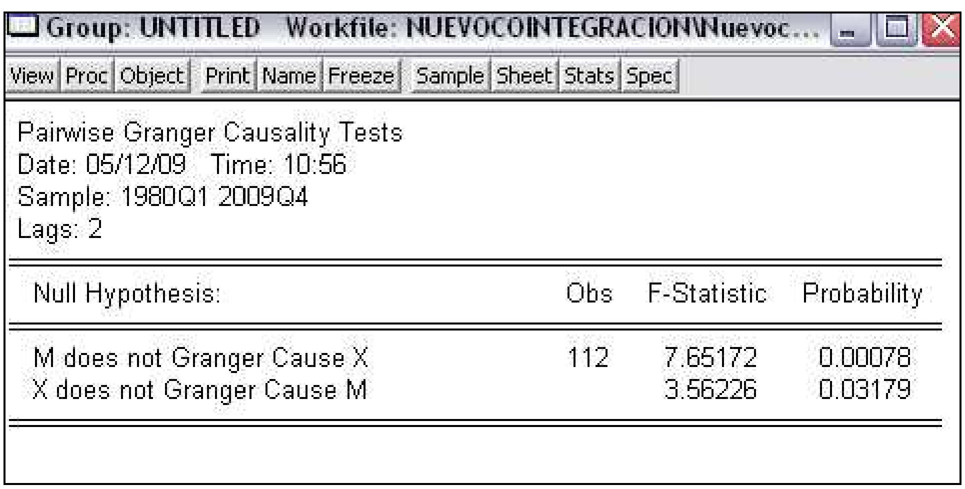
Aparecen los datos de las dos series X y M. Una vez hecho esto en la ventana que despliega vamos a View, a continuación Granger Causality, los pasos se ilustran a continuación.

Cuadro 12.3.4.4.15: Procedimiento para obtener la prueba Granger.



Y en la ventana que despliega, por de fault el programa indica que se utilizaran 2 rezagos y aparece el siguiente cuadro:

Cuadro 12.3.4.4.16: Prueba de Granger.



Group: UNTITLED Workfile: NUEVOCOINTEGRACION\Nuevoc...

View Proc Object Print Name Freeze Sample Sheet Stats Spec

Pairwise Granger Causality Tests
Date: 05/12/09 Time: 10:56
Sample: 1980Q1 2009Q4
Lags: 2

Null Hypothesis:	Obs	F-Statistic	Probability
M does not Granger Cause X	112	7.65172	0.00078
X does not Granger Cause M		3.56226	0.03179

En este Test se establecen las hipótesis nulas de que X no explica a M y a su vez, que M no explica a X. Para aceptarlas necesitamos que la probabilidad asociada (α) debe ser mayor a 5%. Como se observa es menor a 0.05 y concluimos aceptando la hipótesis alterna: X si explica a M y M explica a X. Es decir, X si impacta o causa efectos en M y viceversa, *en el corto plazo*.

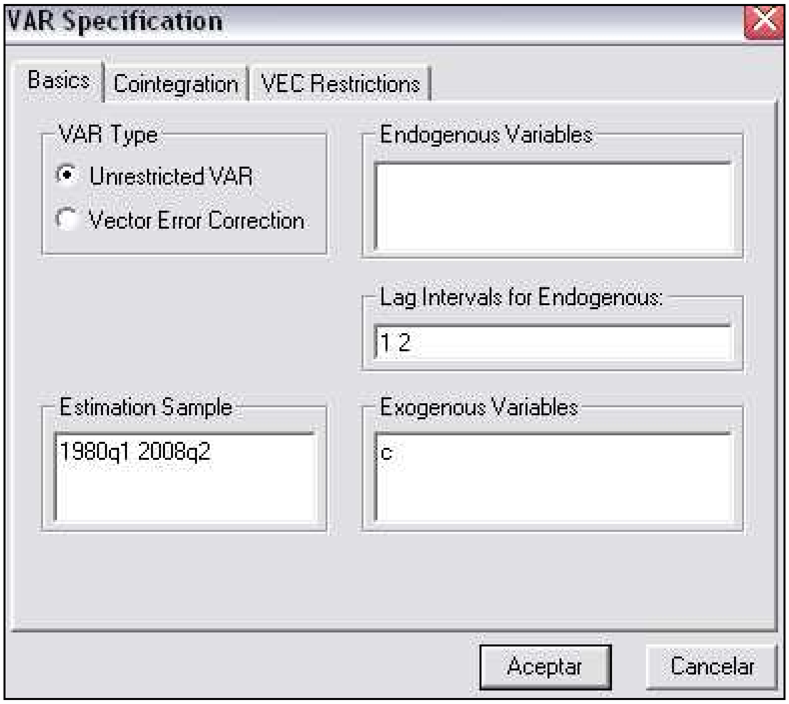
Los resultados anteriores inducen a realizar el análisis de dicha relación, de X y M, en el corto y largo plazo. **Para ello primero se utilizará el modelo VAR para el largo plazo y a partir del cual se obtendrá el VEC para el corto plazo** (*Instituto L.D. Klein citado por UAM en 2004 y* Pérez 2006:635), i.e., él será el referente básico porque expresa la relación reciproca de una variable sobre otra para determinar su valor, relación que se explica con el valor de los coeficientes de las variables X y M endógenas como a continuación se ilustra:

Paso 2.2: Modelo VAR a partir del cual se obtendrá el VEC para el corto plazo.

2.2.1. Con la estimación del modelo VAR se soluciona la necesidad de la especificación estructural de los modelos mediante la modelización de cada variable endógena en el sistema de ecuaciones como función de los valores retardados, de todas las variables endógenas del sistema (Pulido, 1999: 375). Téngase presente que aquí, en la especificación del modelo VAR, se utilizan los valores originales (actuales) de las variables endógenas: X y M. Ello indica que un modelo VAR contiene los propios rezagos de la variable en estudio y los del resto de las variables del sistema de ecuaciones; *no considera ninguna restricción a priori* y como se dice en el marco teórico, todas las variables son identificadas como endógenas. Lo único que se conoce a priori es el número de rezagos (que, se sugiere en general sea alrededor de un tercio del número de términos de la serie, aunque en la práctica es el criterio del investigador el que debe de prevalecer para decir cuántos y porqué) rezagos que se hacen en las variables explicativas que se incluyen en el modelo o sistema de ecuaciones simultáneas.

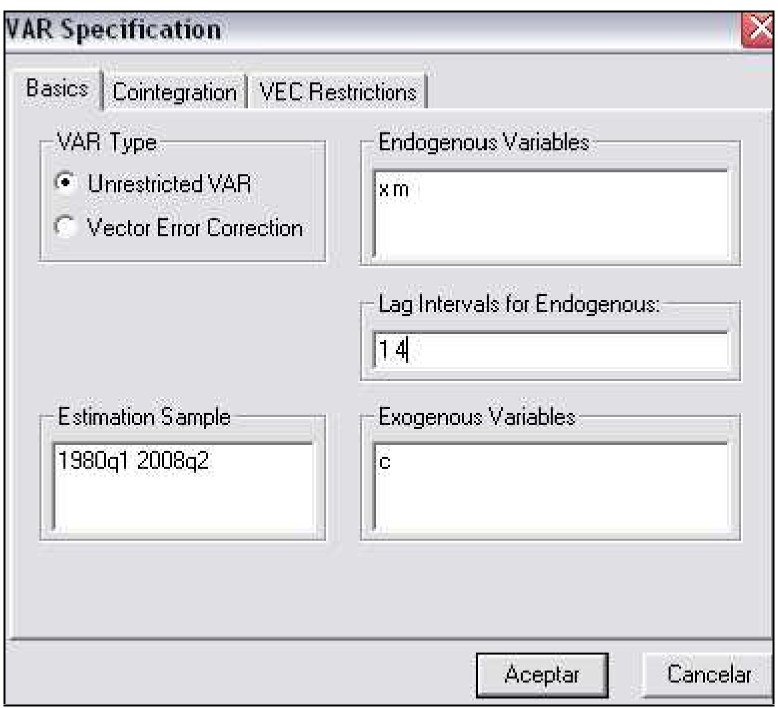
En Eviews para estimar el modelo VAR vamos a Quick, en seguida Estimate VAR y se despliega la ventana siguiente:

Cuadro 12.3.4.4.17: Especificación del Tipo de Modelo VAR en Eviews.



En la cual especificamos en “VAR Type” el que por default marca el programa “Unrestricted VAR”; en “Endogenous Variables” escribimos X_M; en “Lag Invervals for Endogenous” anotamos 1_4 porque así lo deseamos como investigadores; y el resto lo mantenemos igual. Dichas especificaciones se observan en la siguiente ventana:

Cuadro 12.3.4.4.18: Especificación del Tipo de Modelo VAR en Eviews.



Luego en ese mismo cuadro vamos a la pestaña “cointegration”/aceptar, y aparece el cuadro siguiente que contiene el modelo VAR:

Cuadro 12.3.4.4.19: Análisis de Regresión del Modelo VAR.

Vector Autoregression Estimates
Date: 05/11/09 Time: 20:42
Sample(adjusted): 1981:1 2008:2
Included observations: 110 after adjusting Endpoints
Standard errors in ( ) & t-statistics in [ ]

	X	M
X(-1)	1.096537	-0.032654
	(0.16708)	(0.18710)
	[6.56311]	[-0.17453]
X(-2)	-0.004803	0.659480
	(0.25024)	(0.28024)
	[-0.01919]	[2.35328]
X(-3)	-0.471570	-1.154697
	(0.24209)	(0.27110)
	[-1.94793]	[-4.25923]
X(-4)	0.484396	0.754283
	(0.16698)	(0.18699)
	[2.90100]	[4.03383]
M(-1)	-0.394785	0.751834
	(0.14206)	(0.15908)
	[-2.77908]	[4.72605]
M(-2)	0.549528	-0.009394
	(0.20503)	(0.22961)
	[2.68020]	[-0.04091]
M(-3)	-0.023330	0.422376
	(0.21602)	(0.24192)
	[-0.10800]	[1.74596]
M(-4)	-0.197778	-0.351932
	(0.14829)	(0.16606)
	[-1.33373]	[-2.11926]
C	76.17676	277.6340
	(290.536)	(325.359)
	[0.26219]	[0.85331]
R-squared	0.995426	0.994418
Adj. R-squared	0.995064	0.993976
Sum sq. Resids	2.65E+08	3.32E+08
S.E. equation	1619.361	1813.458
F-statistic	2747.624	2249.030
Log likelihood	-964.2650	-976.7174

Akaike AIC	17.69573	17.92213
Schwarz SC	17.91668	18.14308
Mean dependent	30258.28	32443.36
S.D. dependent	23048.87	23364.28
Determinant Residual Covariance		3.12E+12
Log Likelihood (d.f. adjusted)		-1894.501
Akaike Information Criteria		34.77275
Schwarz Criteria		35.21464

Vemos que el cuadro contiene las ecuaciones de la relación entre las variables, ahora rezagadas. El algoritmo para su determinación empezó buscando VAR, mismo cuyo cuadro muestra los coeficientes de cada variable en su ecuación correspondiente, qué, como se indicó, describe la *especificación estructural* del modelo, i.e., la relación de X y M; dicha relación muestra los valores de X en función de sus *cuatro rezagos (* valores omitidos que corresponden a los cuatro trimestres de 1980), como de M también en cuatro rezagos. Lo mismo sucede con la ecuación de M que la muestra en función de si misma y de X rezagadas cuatro veces; además de su ordenada al origen (C) correspondiente.

Reiteramos el modelo VAR al usar rezagos tiene como objetivo principal el identificar el impacto de los valores pasados (se omitieron los cuatro trimestres de 1980 porque están muy lejanos e impactan menos a los valores de 2008:2) en el valor actual de las variables, en este caso X y M. En otras palabras cualquier serie se rezaga con el propósito de detectar el impacto de sus valores de sus últimos periodos en el valor actual de la misma. Así, el coeficiente obtenido para X(-1), que vale 1.096537 expresa mejor la relación del periodo n-1 con n; el coeficiente de X(-3) es -0.471570 y expresa el impacto de X en el periodo n-3 en el valor actual de X. *Ahora bien, en el ámbito de la cointegración es importante decir* que la otra variable actúa como endógena determinando el valor de X: su valor actual está determinado por M rezagada en un número igual de periodos (rezagos) que X. De conformidad con Carrascal, et al (2001): los valores iníciales de X tienen menor impacto en su valor actual, de ahí que para mejorar este último se acostumbre calcular (rezagar/omitir) su valor con respecto a los valores cercanos al actual, ya que es más significativo su impacto que el de los valores iniciales de la serie.

Concluimos diciendo que la relación entre X y M se expresa con los coeficientes de las ecuaciones de cada una de ellas.

Paso 3- : Relación de equilibrio de X y M en el largo plazo:

Al ya conocer el modelo VAR ahora podemos iniciar el análisis de la relación de X y M en el largo plazo, mediante el siguiente procedimiento, que **implica realizar antes** la prueba de Johansen que se muestra en seguida:

Paso 3.1: Realizar el Test de Johansen para conocer el número de ecuaciones de cointegración <u>y verificar si X y M están cointegradas</u>, es decir, si existe relación a largo plazo entre ellas.

Para ello recuerde que econométricamente se dice que las series temporales que son no estacionarias de orden I(1) están cointegradas linealmente si existe una combinación lineal de ellas que mediante la primera diferencia, se conviertan en estacionarias, de orden I(0). El vector de coeficientes que crea esta serie estacionaria (o ecuación de cointegración) es el vector cointegrante.

Así, comencemos diciendo que sabemos que X y M son no estacionarias y que para modelar conjuntamente su efecto en el largo plazo debemos utilizar un modelo VAR. Así, conforme al marco teórico, recuerde que dos variables no estacionarias con tendencia estocástica están cointegradas si se mueven juntas (ejemplo: el ingreso y el gasto) en el tiempo y si sus diferencias son estables (es decir constantes o estacionarias) debido a que gravitan los valores de cada una de las variables en torno a μ y σ^2 constantes.

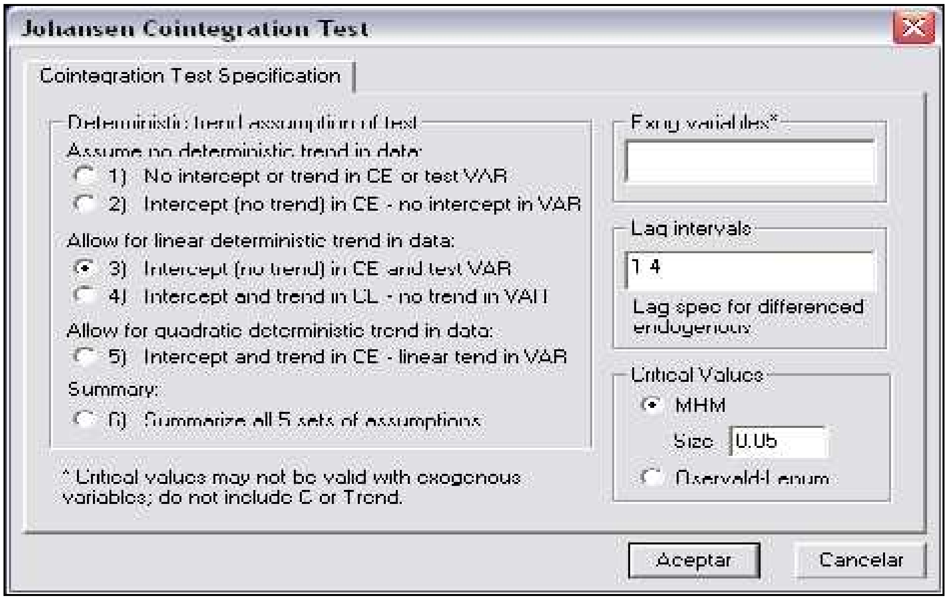
Así, para verificar si X y Y están integradas se empieza con el

Test de Johansen

Para ello iniciemos aplicando <u>la prueba de cointegración de Johansen</u>, la cual, como antes se indicó, además de mostrar si o no existe relación de X con M en el largo plazo, <u>a su vez nos dice el número de ecuaciones de cointegración a utilizar para enseguida predecir los valores de X y M.</u> Así, en la base de datos vamos a Quick/ show: escribimos X M/ok , enseguida,allí, vamos a: view/cointegration test/ y aparece el cuadro de Johansen en el que no cambiamos nada/ok y aparece el cuadro siguiente.T**ambién se puede obtener así:**: Quick/group statistics/ cointegration test y aparece el cuadro de “ series list”, en él escribimos x m /ok/aparece el cuadro de Johansen, en el que no cambiamos ninguna de las datos incluidos por Eviews por default/ok y aparece el siguiente cuadro:

Concretamente, pasemos a ilustrar lo antes dicho ahora identificado si estas variables están cointegradas a través del Test de Johansen. Para ello, vamos el WorkFile y seleccionamos las variables endógenas (X y M, en este orden), dando click derecho las abrimos (Open) como grupo (As Group), lo cual abrirá una ventana con los valores de las variables, tal y como se realizó en ocasiones anteriores. Una vez ahí, vamos a View/Cointegration Test, lo que desplegará una ventana como la siguiente:

Cuadro 12.3.4.4.14: Prueba de Cointegración de Johansen.



Como se dijo, este Test nos permite determinar si las variables están cointegradas y, además, nos proporciona el número de ecuaciones de cointegración. En la vantana anterior hay que especificar el número de retardos (Lag intervals), que para nuestro caso es de 1 a 4. Asimismo, en “Deterministic trend assumption of test” el programa da por default el 3) Intercept (no trend) in CE and test VAR”/Ok, obteniendo:

Cuadro 12.3.4.4.20: Análisis de Resultados de la Prueba de Cointegración de Johansen.

Date: Time:
Sample (adjusted): 1981Q2 2008Q2
Included observations: 109 after adjustments
Trend assumption: Linear deterministic trend
Series: X M
Lags interval (in first differences): 1 to 4

Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.075453	16.18542	15.49471	0.0393
At most 1 *	0.067643	7.634277	3.841466	0.0057

Trace test indicates 2 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None	0.075453	8.551144	14.26460	0.3254
At most 1 *	0.067643	7.634277	3.841466	0.0057

Max-eigenvalue test indicates no cointegration at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

X	M
1.77E-07	-2.27E-07
4.64E-07	-4.25E-07

Unrestricted Adjustment Coefficients (alpha):

D(X)	-338952.3	226596.0
D(M)	-121365.9	414465.7

1 Cointegrating Equation(s): Log likelihood -3366.796

Normalized cointegrating coefficients (standard error in parentheses)

X	**M**
1.000000	**-1.278520**
	(0.11911)

Adjustment coefficients (standard error in parentheses)

D(X)	-0.060088
	(0.02623)
D(M)	-0.021515
	(0.02938)

Como se observa hasta arriba del cuadro anterior, la prueba de Johansen establece las siguientes hipótesis:

Ho: No existe ninguna relación de cointegración de X con M.

Ha: Existe alguna relación de cointegración de X con M.

Para verificar hipótesis nula (Ho) utilizamos la estadística Trace Statistic, si es **mayor** al valor crítico al 5 se rechaza la Ho; si es **menor** se acepta la Ho. Asimismo, el valor de contraste (Eigenvalue) si es mayor a 0.05 se rechaza Ho, y si es menor a 0.05 se acepta la Ho. También, Eviews nos proporciona una probabilidad asociada al estadístico Trace, misma que si es menor a 0.05 rechazamos Ho y si es mayor la aceptamos.

Para nuestro caso probamos primero “None” que significa que no existe ninguna relación de cointegración entre X y M. Así, como el valor del estadístico Trace Statistic (16.18) es mayor al valor Crítico (15.49), aceptamos la Ha, existe cointegración entre X y M. Igualmente como su probabilidad asociada es menor a 0.05, aceptamos Ha, también, porque el valor del estadístico Eigenvalue es mayor a 0.05, se rechaza Ho.

Se hace un planteamiento similar de hipótesis para el siguiente caso **(At most 1**):

Ho: No existe a lo más una relación de cointegración.

Ha: Como máximo existe una relación de cointegración.

Analizando los valore de los estadísticos se arriba a la misma interpretación, es decir, concluimos que se rechaza Ho.

Por tanto, decimos que X y M están cointegradas y que existe una relación de cointegración entre ellas en el largo plazo, que son las ecuaciones que a continuación calcularemos con el Mecanismo de Corrección del Error (MCE o VEC).

Referencias para verificar que X y M están cointegradas:

Prueba 1:

Ho: No existe ninguna relación de cointegración entre X y M.

Ha: Existe alguna relación de cointegración entre X y M.

α = 5%, es la probabilidad de rechazar Ho si es cierta.

Prueba 2:

Ho: No existe cuando más una relación de cointegración entre X y M.

Ha: Como máximo existe una relación de cointegración entre X y M.

α = 5%, es la probabilidad de rechazar Ho si es cierta.

Interpretación de la prueba 1:

1.- Al ser el estadístico "Trace Statistic" (16.18) mayor que el valor crítico (15.49) y como su probabilidad asociada es menor al 5% (que es 3.93%) se rechaza la Ho, por tanto, existe relación lineal de cointegración entre ambas series. Asimismo, al ser el valor de contraste (Eigenvalue: 0.075453) **mayo**r que 0.05 se rechaza la Ho.

Interpretación de la prueba 2:

1.- Al ser el estadístico "Trace Statistic" (7.6334277) mayor que el valor crítico (3.841466) al 5% se rechaza la Ho, por tanto, existe cointegración entre X y M. Asimismo, el valor de contraste (Eigenvalue: 0.067643) al ser mayor que 0.05 se rechaza la Ho.

Comentario adicional: al final del cuadro, hasta abajo, se indica el número de ecuaciones de cointegración, con letras sombreadas en negritas; se ve que es una sólo ecuación de cointegración.

Conclusiones:

1.- X y M <u>tienen una relación lineal o equilibrio de largo plazo al estar cointegradas.</u> La cointegración expresa la presencia de un **equilibrio** a largo plazo hacia el que converge el sistema económico a largo plazo. X y M tienen una relación lineal o equilibrio a largo plazo al estar cointegradas, <u>cuya ecuación se obtendrá enseguida;</u>

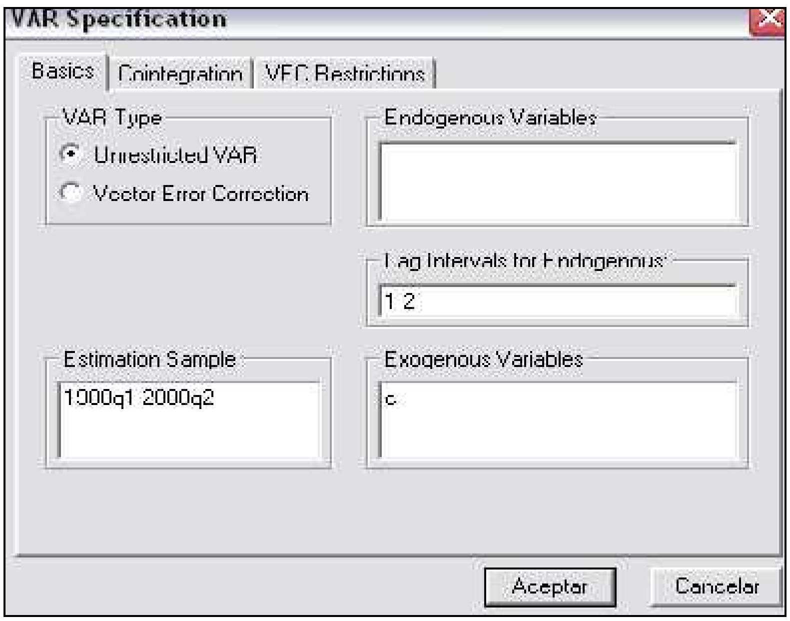
2.- Dicho equilibrio se obtiene con la ecuación de un modelo VAR, mismo que establece o expresa la relación a largo plazo de X y M y cuya característica es que siendo X y M no estacionarias las convierte en estacionarias (diferenciándolas una vez), por lo que la predicción resulta más segura en el largo plazo al mostrar sus datos un comportamiento estacionario: sus valores gravitan en torno a μ y σ^2 constantes; <u>tal que si se alejan gradualmente se corrige ese alejamiento a partir del modelo VAR con el Mecanismo de Corrección del Error, MEC</u> y por consiguiente lo anterior indica que las *diferencias* (denominadas "errores") que se presentan en las dos variables en su devenir en el tiempo, en la ecuación de cointegración se interpretan como el " error de desequilibrio" (error Correction) para cada punto particular en el tiempo. Este error o *diferencia* se corrige con el modelo VEC, cuya *ecuación* es la referencia básica y que se obtiene con el siguiente procedimiento:

Corrección del error con el mecanismo del modelo VEC

Una vez que hemos comprobado que X y M están cointegradas (Pulido et al, 1999: 400) de orden 1, ahora ya podemos estimar un modelo de corrección de error (VEC), que como se indicó en el marco teórico el error se origina en las desviaciones de algunos de los valores de la serie con respecto a su curva con comportamiento constante en su media y desviación estándar (es decir, se elimina o se reduce la tendencia, creciente o decreciente, de la curva que daba lugar a una serie no estacionaria), con base en los siguientes pasos que tienen como operación fundamental la diferenciación de sus valores, con objeto de eliminar la autocorrelación entre los términos de la serie y homogeneizar su valor en el tiempo.

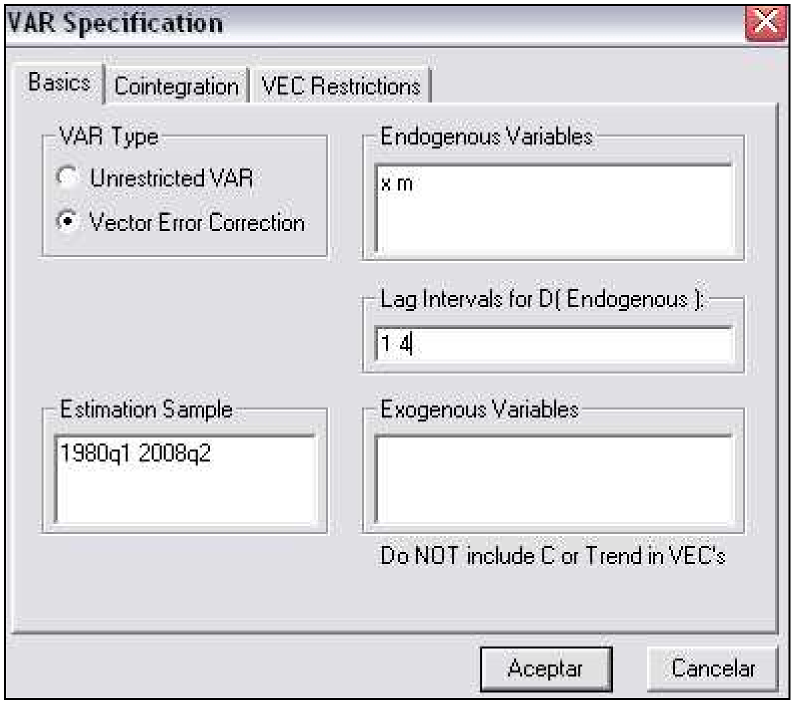
Así, de conformidad con Pérez (2006:635) para estimar el modelo VEC en el menú principal vamos a: Quick/Estimate VAR/Ok. Como se hizo anteriormente, y nos desplegará una ventana como la siguiente:

Cuadro 12.3.4.4.21: Especificación del Tipo de Modelo VEC en Eviews.



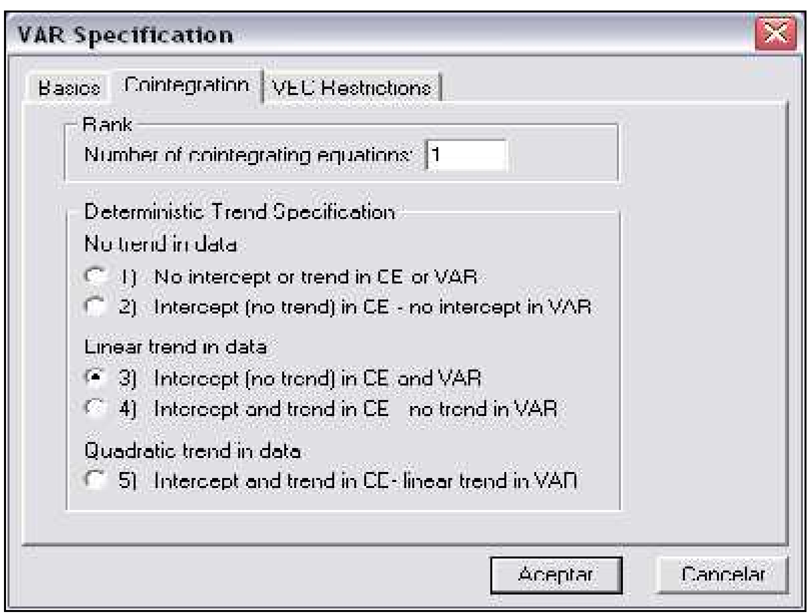
En la especificación, hasta abajo está la ventana: Estimation Sample, allí establecemos que queremos la estimación desde el primer trimestre de 1980 al segundo de 2008; luego en VAR Type seleccionamos “Vector Error Correction”; y en “Endogenous Variables” escribimos X_M, con los retardos 1 y 4 (lags intervals for Endogenous) que nos da Eviews por default, con lo cual el cuadro anterior ahora presenta la siguiente estructura

Cuadro 12.3.4.4.22: Especificación del Tipo de Modelo VEC en Eviews.



Luego en ese mismo cuadro vamos a la **pestaña de Cointegration/clic**, misma que ahora visualiza el siguiente cuadro:

Cuadro 12.3.4.4.23: Prueba de Cointegración para el Modelo VEC.



En este nuevo cuadro, se muestra hasta arriba que existe una sola ecuación de cointegración. Al respecto, *conviene abundar sobre este punto diciendo de manera* *reiterada que el máximo número de ecuaciones de cointegración que se puede manejar es el número de variables endógenas menos uno*; en nuestro caso es 2-1= una ecuación de cointegración). Lo anterior en automático Eviews lo indica en "Rank: Number of cointegrating equations", es 1. Y en "Deterministic Trend Specification", seleccionamos "Linear trend in data" y en ella especificamos "intercep (no trend) in CE and VAR". El resto permanece en blanco/Ok., lo cual generará el siguiente cuadro que contiene la composición de la **ecuación de cointegración,** misma que indica la relación de cointegración normalizada asumiendo una relación de cointegración con ella misma de 1.

Cuadro 12.3.4.4.24: Análisis de Regresión del Modelo VEC.

Vector Error Correction Estimates
Date: 05/11/09 Time: 20:47
Sample(adjusted): 1981:2 2008:2
Included observations: 109 after adjusting Endpoints
Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
X(-1)	1.000000
M(-1)	-1.278801
	(0.11919)
	[-10.7292]
C	11074.14

Error Correction:	D(X)	D(M)
CointEq1	-0.060054	-0.021518
	(0.02620)	(0.02936)
	[-2.29170]	[-0.73287]
D(X(-1))	0.163945	0.003842
	(0.15567)	(0.17442)
	[1.05314]	[0.02203]
D(X(-2))	0.169613	0.593994
	(0.16188)	(0.18137)
	[1.04778]	[3.27500]
D(X(-3))	-0.337486	-0.463862
	(0.16763)	(0.18782)
	[-2.01326]	[-2.46974]
D(X(-4))	0.597677	0.632300
	(0.16413)	(0.18389)
	[3.64157]	[3.43846]
D(M(-1))	-0.290610	-0.042168
	(0.13834)	(0.15500)
	[-2.10074]	[-0.27206]

D(M(-2))	0.015439	-0.296073
	(0.14194)	(0.15903)
	[0.10877]	[-1.86175]
D(M(-3))	0.115373	0.213223
	(0.14327)	(0.16053)
	[0.80526]	[1.32826]
D(M(-4))	-0.164483	-0.024748
	(0.14225)	(0.15939)
	[-1.15626]	[-0.15527]
C	580.7010	331.3434
	(213.703)	(239.437)
	[2.71733]	[1.38384]
R-squared	0.525973	0.532119
Adj. R-squared	0.482880	0.489584
Sum sq. Resids	2.36E+08	2.96E+08
S.E. equation	1544.472	1730.458
F-statistic	12.20544	12.51024
Log likelihood	-949.7455	-962.1393
Akaike AIC	17.61001	17.83742
Schwarz SC	17.85692	18.08433
Mean dependent	770.0642	764.8532
S.D. dependent	2147.752	2422.139
Determinant Residual Covariance		2.80E+12
Log Likelihood		-1860.897
Log Likelihood (d.f. adjusted)		-1871.386
Akaike Information Criteria		34.74102
Schwarz Criteria		35.28423

Vemos que además de la ecuación de cointegración (arriba), enseguida se observa que hay dos ecuaciones y por consiguiente decimos que sus *coeficientes muestran la relación a largo plazo de X con M* y que es con ellos con que se corrige el error, además de que se usan para predecir los valores futuros de esas variables; dichas ecuaciones para usarse primero se rezagan y luego se les aplica "la primera diferencia".

Por otra parte, en lo que se refiere al grado de relación que existe entre las dos variables (dentro de cada ecuación, dado que son endógenas) expresado, por R^2 , vemos que a diferencia del modelo VAR donde esta estadística refleja valores muy altos de asociación mutua (alrededor del 99%), aquí con el VEC dicha relación es mucho más baja (52.9 % en la primera ecuación y 53.2 % en la segunda), sobre todo en lo que se refiere a la R^2 ajustada (48.2% y 48.5% respectivamente). Ello en opinión de Pulido et al no debe sorprendernos mucho. ¿ A qué se debe? ¿ Cómo se interpretan en VAR y en VEC los R^2 ?

Aquí los valores bajos de estos dos coeficientes indican que no hay autocorrelación entre sus términos. De igual manera, al no haber *autocorrelación* es un indicio de que con VEC ésta se eliminó por medio de la "primera diferencia" aplicada a los valores rezagados; esta autocorrelación no se había eliminado con VAR puesto que nada más los rezagó. Esto es muy importante porque ahora los valores proyectados estarán "limpios", libres de toda contaminación cronológica y serán estacionarios, i.e., cuando se alejen de la media, siempre serán corregidos con el VEC y tenderán a gravitar a su alrededor, tendrán, por consiguiente, una σ^2 y covarianza constantes.

Así, continuando con el análisis de la ecuación de cointegración, hasta arriba del cuadro, ella muestra la primera diferencia (D) para X, "D(X)" y para M, "D(M)" , la cual indica la corección del error en cada variable, es decir β, por ejemplo, es 0.060054 para D(X) y es 0.021518 para D(M), que es el cambio que se realiza entre D(M) y X.

Para X en D(X-1) el coeficiente de la variable dependiente D(X) que ya contiene un rezago indica los cambios que se registran entre X (en su valor actual) y ella misma con un rezago, que en este caso es 0.163945. Se observa que a medida que aumentan los rezagos los cambios que se operan entre X y sus valores "pasados" (X-2), X(-3) y X(-4) tienden a ser mayores , i.e., los errores corregidos (coeficientes) aumentan de valor y eso se debe a la mayor lejanía de X(-2), X(-3) y X-4) del valor actual de X, situación que acrecienta la corrección del error. Cabe decir que lo mismo se observa e interpreta en M.

Al respecto, tomando como referencia la ecuación de cointegración descrita previamente sólo con literales, todo indica (Rojas, 2009) que **a_1 y a_2, expresan** la velocidad con que se ajusta la desviación incurrida por las variables en el devenir entre el corto y el largo plazo, mismas que toman los valores - 0.06005409622 en la ecuación de las exportaciones y -0.02151755148 en la ecuación de las importaciones, en tanto que **β** es igual a -1.278801088 en ambas. Con esa referencia y dado que la ecuación de cointegración va en las dos ecuaciones (de X y M), **vemos, reiterando, que β**, es el termino de corrección del error en ambas. Lo anterior es importante si se recuerda que β toma el valor de cero en el equilibrio de largo plazo, pero si X y M se desvían de ese equilibrio, entonces β es diferente de 0, como es el caso, por lo que cada una de esas variables se ajusta parcialmente en el corto plazo para restablecer el equilibrio de largo plazo.

Derivado de lo anterior es que puede decirse que *a diferencia de un modelo VAR, con el VEC se realizan ajustes en el corto plazo que mejoran las proyecciones a largo plazo porque la relación entre las variables se vuelve estacionaria tal que los datos proyectados muestran un comportamiento estable en torno a μ y σ^2 constantes.* Luego entonces enseguida se procede a la obtención de las ecuaciones de X y M que nos permitirán proyectar sus valores a futuro.

Obtención de las ecuaciones de X y M para predecir

Dado que el cuadro anterior indica que la ecuación de cointegración es:

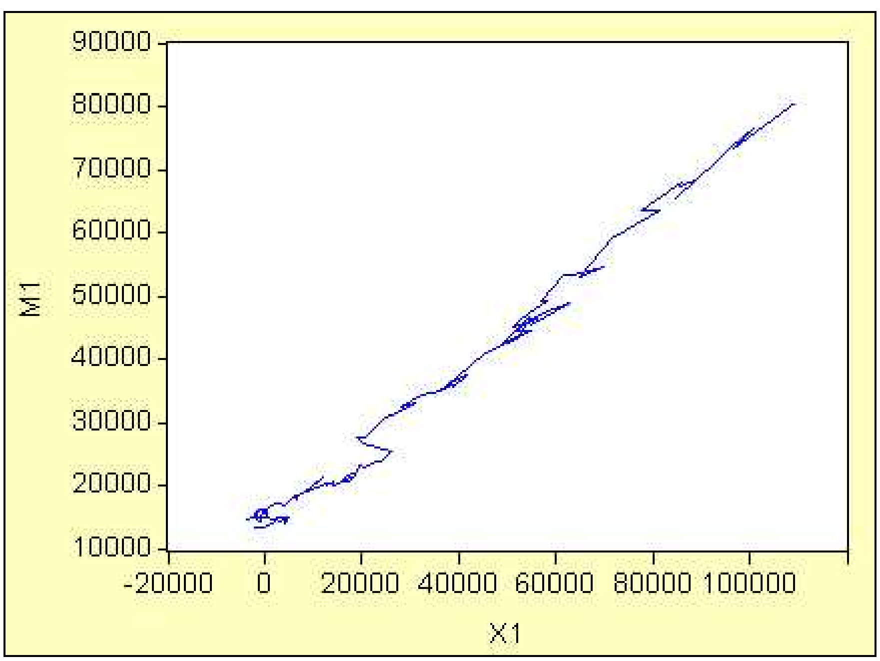
X -1.278801M + 11074.14; entonces para construir su gráfica debemos igualar a cero la ecuación y despejar X y M, respectivamente, a partir de las cuales se construye la gráfica.

X = - 11074.14 + 1.278801M… (1)

M = 8659.78 + 0.781982X… (2)

A cada una de ellas le vamos a dar valores, por ejemplo, para la ecuación 1 le sustituimos los valores originales de M, y para la ecuación 2, los valores de X. Para ello vamos a utilizar Excel para crear las series o variables relacionadas, mismas que pasaremos a Eviews para su posterior graficación, a la ecuación uno llamémosle X1 y a la ecuación 2, M1. Una vez generadas, en el Workfile seleccionamos el grupo en donde ambas están capturadas o en su caso seleccionamos una a una y vamos a Quick/Graph/XY line/, y en el cuadro que despliega decimos que Ok., con lo cual obtendremos la gráfica de relación de cointegración ente X1 y M1:

Gráfica 12.3.4.4.3: Relación de Cointegración entre X1 y M1.



Como puede observarse, la curva tiene pendiente positiva lo cual corrobora la teoría de que al aumentar el valor de una variable (X1) su valor influye de manera positiva en la otra (M1), es decir, que de conformidad con los registros de INEGI se detecta que para poder exportar siempre ha sido necesario importar (insumos, materias primar, maquinaria y equipo, etc.), esto se corrobora de manera contundente en el caso de la Industria Maquiladora, que importa prácticamente todo, que por ejemplo en 2006 fueron responsables del 45% de las exportaciones del país.

Analizando en detalle dicho cuadro (siguiente) referido al modelo VEC, es importante indicar que la estimación del modelo VEC comienza en primer lugar, en opinión de Pulido et al, determinando una o más ecuaciones de cointegración utilizando el método de Johansen. Así, se estiman regresiones de la primera diferencia (D) de cada variable endógena sobre la ecuación de cointegración y las primeras diferencias retardadas de todas las variables endógenas incluidas en el sistema, **cuyo número el investigador lo determina (en este caso de 1 a 4)**.

Interpretación de los coeficientes. Si partimos de la ecuación de cointegración (hasta arriba) en error de corrección vemos “CointEq1” que es la ecuación anterior pero ahora en primera diferencia para X, “D(X)” y M, “D(M)”. Lo cual muestra la corrección del error en cada una de ellas, es decir, en el caso de D(X), vemos -0.06, el cual, suponiendo que D(X) es la variable dependiente, se interpreta como los cambios que se operan entre D(X) y M y no las relaciones entre los niveles (como en el VAR). En el caso de D(X(-1)) el coeficiente de la variable dependiente D(X) se interpreta como el cambio en las diferencias que existen entre D(X) pero rezagada un periodo. Se dice que ahora, al estar los coeficientes de las ecuaciones expresados en diferencias, ellas dejan de ser no estacionarias y ya son susceptibles de ser utilizadas para predecir los valores de X y M.

Ahora bien conviene decir que partiendo del cuadro anterior la **corrección (conciliar el corto con el largo plazo)** del error sigue el siguiente proceso: la variable original, X (que también puede ser M) es sometida en primer lugar a un rezago, en seguida es diferenciada una vez y rezagada 1, 2, 3 y 4 veces, misma que al estar cointegrada con M, ésta también es sometida al mismo proceso e incluida en la misma ecuación (D(X). Similarmente para M se sigue el mismo procedimiento de X. Ello se debe a que necesitamos dos ecuaciones para predecir ambas variables.

Para visualizar las ecuaciones, estando en la ventana del VEC vamos a View/Representations/Ok., lo cual despliega la siguiente ventana:

Cuadro 12.3.4.4.25: Ecuaciones del Modelo VEC.

```
Estimation Proc:
===============================
EC(C,1) 1 4 X M

VAR Model:
===============================
D(X) = A(1,1)*(B(1,1)*X(-1) + B(1,2)*M(-1) + B(1,3)) + C(1,1)*D(X(-1)) + C(1,2)*D(X(-2)) + C(1,3)
*D(X(-3)) + C(1,4)*D(X(-4)) + C(1,5)*D(M(-1)) + C(1,6)*D(M(-2)) + C(1,7)*D(M(-3)) + C(1,8)*D(M(-
4)) + C(1,9)

D(M) = A(2,1)*(B(1,1)*X(-1) + B(1,2)*M(-1) + B(1,3)) + C(2,1)*D(X(-1)) + C(2,2)*D(X(-2)) + C(2,3)
*D(X(-3)) + C(2,4)*D(X(-4)) + C(2,5)*D(M(-1)) + C(2,6)*D(M(-2)) + C(2,7)*D(M(-3)) + C(2,8)*D(M(-
4)) + C(2,9)

VAR Model - Substituted Coefficients:
===============================
D(X) =  - 0.06005409622*( X(-1) - 1.278801088*M(-1) + 11074.13752 ) + 0.1639451205*D(X(-1))
+ 0.16961291*D(X(-2)) - 0.337485587*D(X(-3)) + 0.5976770003*D(X(-4)) - 0.2906096476*D(M(-
1)) + 0.01543856127*D(M(-2)) + 0.1153728258*D(M(-3)) - 0.1644833507*D(M(-4)) +
580.7010363

D(M) =  - 0.02151755148*( X(-1) - 1.278801088*M(-1) + 11074.13752 ) + 0.003842353517*D(X
(-1)) + 0.5939942495*D(X(-2)) - 0.4638615266*D(X(-3)) + 0.6323003611*D(X(-4)) -
0.04216826772*D(M(-1)) - 0.2960731768*D(M(-2)) + 0.2132232439*D(M(-3)) - 0.02474805851
*D(M(-4)) + 331.3434169
```

Obsérvese que los procesos de las ecuaciones para D(X) y D(M) primero se describen con literales (VAR Model) y luego con los valores de esas literales (VAR Model – Substituted Coefficients).

¿**Qué ecuaciones, coeficientes, etc., muestran en detalle el proceso de corrección del error?**

Para contestar estas preguntas establecemos la siguiente **referencia**:

El cuadro anterior generado con Eviews para el modelo VEC, muestra hasta arriba que la ecuación de cointegración es:

X(-1)=1.00

M(-1)=-1.278801 y que C=11074.14

De manera que al generar las ecuaciones D(X) y D(M) vemos que cada una de ellas está constituida por los siguientes términos:

1.- Para D(X):

D(X) = - 0.06005409622*(X(-1) - 1.278801088*M(-1) + 11074.13752) + 0.1639451205*D(X(-1)) + 0.16961291*D(X(-2)) - 0.337485587*D(X(-3)) + 0.5976770003*D(X(-4)) - 0.2906096476*D(M(-1)) + 0.01543856127*D(M(-2)) + 0.1153728258*D(M(-3)) - 0.1644833507*D(M(-4)) + 580.7010363

Observaciones:

i).- Aquí el corrector del error β=-0.06005409622;

ii).- a_1= X(-1)-1.278801088*M(-1);

iii).- a_2 = +11074.13752;

iv).- Los términos de i,ii, iii, producen el equilibrio a corto plazo a la velocidad indicada por los valores de a_1 y a_2 reduciendo o corrigiendo las desviaciones (errores) de los términos de la serie X con respecto a µ, usando el corrector del error β, con objeto de que dichos términos tiendan hacia µ de dicha serie de X, es decir: - 0.06005409622*(X(-1) - 1.278801088*M(-1) + 11074.13752).

v).- Derivado de lo anterior decimos que el resto de los términos de D(X) producen el equilibrio a largo plazo, mismo que se logra rezagando y diferenciado los términos de X, es decir: + 0.1639451205*D(X(-1)) + 0.16961291*D(X(-2)) - 0.337485587*D(X(-3)) + 0.5976770003*D(X(-4)) - 0.2906096476*D(M(-1)) + 0.01543856127*D(M(-2)) + 0.1153728258*D(M(-3)) - 0.1644833507*D(M(-4)) + 580.7010363.

2.- Para D(M):

D(M) = - 0.02151755148*(X(-1) - 1.278801088*M(-1) + 11074.13752) + 0.003842353517*D(X(-1)) + 0.5939942495*D(X(-2)) - 0.4638615266*D(X(-3)) + 0.6323003611*D(X(-4)) - 0.04216826772*D(M(-1)) - 0.2960731768*D(M(-2)) + 0.2132232439*D(M(-3)) - 0.02474805851*D(M(-4)) + 331.3434169

Observaciones:

i).- Aquí el corrector del error β=-0.02151755148;

ii).- a_1= X(-1)-1.278801088*M(-1);

iii).- a_2 = +11074.13752;

iv).- Los términos de i,ii, iii, producen *el equilibrio a corto plazo* a la velocidad indicada por los valores de a_1 y a_2 reduciendo o corrigiendo las desviaciones (errores) de los términos de la serie M con respecto a µ, usando el corrector del error β, con objeto de que dichos términos tiendan hacia μ de dicha serie de X, es decir: - 0.02151755148*(X(-1) - 1.278801088*M(-1) + 11074.13752) .

v).- Derivado de lo anterior decimos que el resto de los términos de D(M) producen el *equilibrio a largo plazo*, mismo que se logra rezagando y diferenciado los términos de M, es decir: + 0.003842353517*D(X(-1)) + 0.5939942495*D(X(-2)) - 0.4638615266*D(X(-3)) + 0.6323003611*D(X(-4)) - 0.04216826772*D(M(-1)) - 0.2960731768*D(M(-2)) + 0.2132232439*D(M(-3)) - 0.02474805851*D(M(-4)) + 331.3434169

vi).- Observe que en D(X) y D(M) los valores de a_1 y a_2 son los mismos, no así el de β.

Interpretación:

1. Las ecuaciones se presentan en una doble modalidad: primero, con literales y segundo, con sus valores correspondientes.
2. Con base en el cuadro anterior, el primer número expresa la pendiente de X con un rezago y en primera diferencia (-0.06005409) multiplicado por la ecuación de cointegración: X con un rezago (X(-1)) menos la pendiente de M con un rezago (-1.278801088) más el valor de la ordenada al origen (11074.13752) + …+ la ordenada al origen de la ecuación de predicción de X (580.7010363).
3. Algo similar se hace en lo que se refiere en la ecuación de M.
4. Estas dos ecuaciones son las que se utilizan para predecir X y M respectivamente.

En el cuadro arriba explicado viene el proceso que muestra que para X fue necesario ir determinando coeficientes de ajuste con base en diferenciaciones y en rezagos. Ejemplo, trabajando para X a través del Vector de Corrección del Error primero diferencia una vez la variable original, en este caso X, y a partir de las indicaciones dadas (de 1 a 4 rezagos) va a incluir en la ecuación de regresión la variable X diferenciada y rezagada 1, 2, 3 y 4 periodos respectivamente.

¿Cómo, dónde se vio la corrección del error, es decir, la bondad del ajuste?

Para ilustrar la corrección del error se comparan la suma de los residuos al cuadrado (Sum sq. Resid) y el error estándar de la ecuación (S.E. Equation) del VAR y el VEC. Para verificar que se corrigió el error, en el VEC dichas estadísticas deben ser menores que en VAR. Lo cual se constata con lo siguiente:

Cuadro 12.3.4.4.26: Comparación entre la suma de los errores al cuadrado y el error estándar de la ecuación del Modelo VAR y VEC.

	PARA X		PARA M	
Modelo	**VAR**	**VEC**	**VAR**	**VEC**
Sum sq. Resid	2.65×10^{8}	2.36×10^{8}	3.32×10^{8}	2.96×10^{8}
S.E. Equation	1619.36	1544.472	1813.458	1730.458

Al respecto es importante decir que los valores en VEC están expresando las diferencias en tanto que en VAR son valores originales.

Se reitera que VEC a diferencia de VAR incluye en cada ecuación de predicción (X o M) la ecuación de cointegración. Además, en el VAR tenemos el valor de los coeficientes en niveles rezagados, en tanto que en el VEC están en primeras diferencias rezagados. Lo anterior se visualiza con la siguiente gráfica que expresa la relación de cointegración de X con M y viceversa. Para obtenerla, estando en la ventana del VEC vamos a View/Cointegration Graph/Ok.

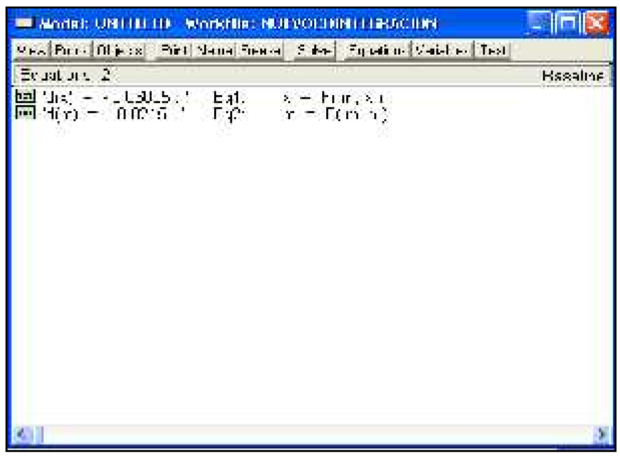
Gráfica 12.3.4.4.4: Relación de Cointegración entre X y M del Modelo VEC.

Se observa en la gráfica que los valores diferenciados giran en torno a una media a largo plazo: ya presentan crecimientos y decrecimientos en el largo plazo como cuando las series X y M eran no estacionarias, *pero ahora corregidas* las desviaciones de sus valores que gravitan en torno a la media aritmética, μ.

Paso 4: Predicción.

Una vez generadas las ecuaciones procedemos a realizar la predicción, para ellos debemos modificar el Rango (Range) en el Workfile. Así, dando doble click sobre el Range se abre una ventana en la cual especificaremos que será hasta 2009Q4. Une vez hecho esto, enseguida vamos a: Quick/estímate VAR/escribo “VAR TYPE:VEC, Endogeneous: X_ M/ok, lags: 1_4; enseguida aparece Cuadro VEC donde seleccionamos: Procs/Make Model/Ok., apareciendo la siguiente ventana:

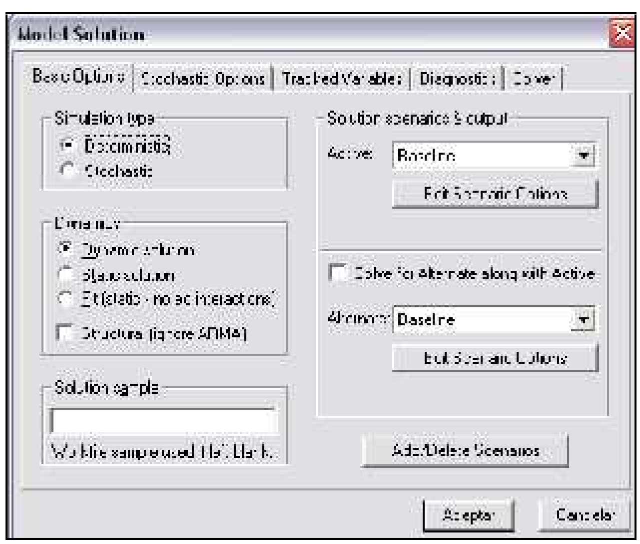
Cuadro 12.3.4.4.27: Pantalla del Workfile.



Aquí aparecen las dos ecuaciones que necesitamos para predecir: X y M.

En esta nueva ventana presiónanos el botón “Solve”, mismo que desplegará un cuadro como el siguiente:

Cuadro 12.3.4.4.28: Solución del modelo en Eviews.



En el cual especificamos en “Solution Sample”: 2008q3_2009q4, dejando igual el resto de opciones/Ok., lo que generará dos nuevas series en el Workfile, mismas que por default Eviews nombra como X_0 y M_0 que son la predicción a partir de 2008q3, lo cual se observa en la siguiente tabla, que se obtiene modificando el Sample, ahora a 2009q4:

Cuadro 12.3.4.4.29: Modificación del rango.

Obs	X	X_0	M	M_0
1980:1	5811	5811	7146	7146
1980:2	6060	6060	8614	8614
1980:3	6263	6263	9442	9442
1980:4	6726	6726	10092	10092
1981:1	8156	8156	10741	10741
1981:2	8132	8132	12010	12010
1981:3	6989	6989	12067	12067
1981:4	7834	7834	12534	12534
1982:1	7018	7018	11112	11112
1982:2	7709	7709	10232	10232
1982:3	8038	8038	8413	8413
1982:4	8207	8207	7105	7105
1983:1	7547	7547	5908	5908
1983:2	8116	8116	6707	6707
1983:3	8270	8270	7236	7236
1983:4	8996	8996	7218	7218
1984:1	9769	9769	7595	7595
1984:2	9502	9502	8258	8258
1984:3	9288	9288	8997	8997
1984:4	9271	9271	8797	8797
1985:1	9245	9245	9071	9071
1985:2	8379	8379	8901	8901
1985:3	9110	9110	8796	8796
1985:4	9124	9124	8291	8291
1986:1	7517	7517	7917	7917
1986:2	7342	7342	8194	8194
1986:3	7025	7025	7659	7659
1986:4	8043	8043	7531	7531
1987:1	8720	8720	7342	7342
1987:2	9549	9549	8026	8026
1987:3	9403	9403	8872	8872
1987:4	9697	9697	8890	8890
1988:1	10339	10339	9580	9580
1988:2	10749	10749	10713	10713
1988:3	10343	10343	11797	11797
1988:4	10664	10664	12382	12382
1989:1	11697	11697	12727	12727
1989:2	12347	12347	13618	13618
1989:3	11708	11708	13770	13770
1989:4	12351	12351	13810	13810
1990:1	13122	13122	15313	15313
1990:2	12469	12469	14128	14128

1990:3	14204	14204	15982	15982
1990:4	16276	16276	18100	18100
1991:1	13391	13391	15534	15534
1991:2	14807	14807	18659	18659
1991:3	14606	14606	18764	18764
1991:4	15283	15283	19777	19777
1992:1	14460	14460	19612	19612
1992:2	15467	15467	21365	21365
1992:3	15484	15484	22339	22339
1992:4	16257	16257	22792	22792
1993:1	15628	15628	21289	21289
1993:2	16952	16952	22597	22597
1993:3	16683	16683	23349	23349
1993:4	18489	18489	23916	23916
1994:1	18062	18062	24843	24843
1994:2	19406	19406	26882	26882
1994:3	19461	19461	27369	27369
1994:4	21443	21443	28939	28939
1995:1	23017	23017	24372	24372
1995:2	24056	24056	23700	23700
1995:3	24413	24413	24864	24864
1995:4	25543	25543	25671	25671
1996:1	26626	26626	26790	26790
1996:2	28417	28417	28162	28162
1996:3	29104	29104	29969	29969
1996:4	31169	31169	32904	32904
1997:1	30415	30415	30635	30635
1997:2	32659	32659	33843	33843
1997:3	33334	33334	35924	35924
1997:4	34910	34910	38581	38581
1998:1	33704	33704	36983	36983
1998:2	35476	35476	38868	38868
1998:3	34207	34207	38921	38921
1998:4	36762	36762	41369	41369
1999:1	35424	35424	39062	39062
1999:2	39268	39268	42134	42134
1999:3	40821	40821	43969	43969
1999:4	43397	43397	47694	47694
2000:1	44326	44326	49256	49256
2000:2	47463	47463	50949	50949
2000:3	49489	49489	53512	53512
2000:4	51598	51598	57843	57843
2001:1	46942	46942	51718	51718
2001:2	47806	47806	51382	51382
2001:3	45788	45788	49111	49111
2001:4	45630	45630	51653	51653
2002:1	43356	43356	47115	47115
2002:2	48427	48427	51287	51287
2002:3	48090	48090	51317	51317
2002:4	48267	48267	52558	52558
2003:1	46400	46400	48711	48711
2003:2	47911	47911	49708	49708

2003:3	49427	49427	51270	51270
2003:4	51653	51653	54274	54274
2004:1	51683	51683	53170	53170
2004:2	57264	57264	56798	56798
2004:3	57071	57071	58313	58313
2004:4	58886	58886	63218	63218
2005:1	56750	56750	59310	59310
2005:2	64774	64774	64869	64869
2005:3	64802	64802	65271	65271
2005:4	70192	70192	72275	72275
2006:1	69935	69935	69407	69407
2006:2	75941	75941	75804	75804
2006:3	74990	74990	75685	75685
2006:4	76070	76070	78271	78271
2007:1	72625	72625	74984	74984
2007:2	79966	79966	81477	81477
2007:3	82752	82752	83751	83751
2007:4	86569	86569	87513	87513
2008:1	82609	82609	84099	84099
2008:2	92093	92093	94110	94110
2008:3		92063.61		93820.8
2008:4		98103.14		100625.4
2009:1		95007.31		96496.7
2009:2		102448.1		104636.7
2009:3		101289.4		103111.3
2009:4		107740.8		110137.9

Comentarios técnicos: La predicción realizada con la técnica de la cointegración se basa o usa los valores históricos de las variables que integran el modelo o sistema de ecuaciones simultáneas, ahora corregido y garantizando su relación en el largo plazo. Se observa que *como en la técnica de Box & Jenkins*, usa los valores pasados pero ahora de dos variables: X y M, para predecir; mismas que al cointegrarse, se retroalimentan y determinan conjuntamente, como variables endógenas, para determinar sus valores futuros con mayor consistencia porque giran alrededor de su media, garantizando así su relación de equilibrio en el largo plazo.

Comentarios económicos: Al comparar los valores futuros de estas dos variables clave del comercio exterior mexicano, se observa que en general seguirán aumentando las importaciones para aumentar las exportaciones, pero se observa que en el primer trimestre de 2009 ambas decrecen, situación similar al comportamiento real de la economía debido a la crisis financiera que se inició en octubre de 2008 y que un no toca fondo. Será interesante observar qué pasa en el tercero y cuarto trimestres de 2009, ¿Realmente caerán las exportaciones en 2009/3 y aumentarán en 2009/4 como lo visualiza este ejercicio econométrico con VEC?

A: Análisis comparativo de la predicción realizada con econometría tradicional:

Desgraciadamente no se puede predecir para los siguientes periodos con la econometría tradicional en virtud de que con Eviews se necesitan conocer o dar valores a las Importaciones (M) suponiendo que el modelo a estimar sea X = f(M) y viceversa, es decir, se necesitan conocer o dar valores hipotéticos a la variable X si el modelo a estimar fuera M = f(X).

Sin embargo, si se pueden comparar las "estimaciones" tanto de las X como de las M obtenidas con la econometría tradicional y con el método MCE. Así, por ejemplo, en Eviews vamos a Quick/Estimate Equation/, en la ventana que aparece escribimos X_c_M/Ok., se obtienen los siguientes resultados:

Cuadro 12.3.4.4.30: Análisis de Regresión Lineal.

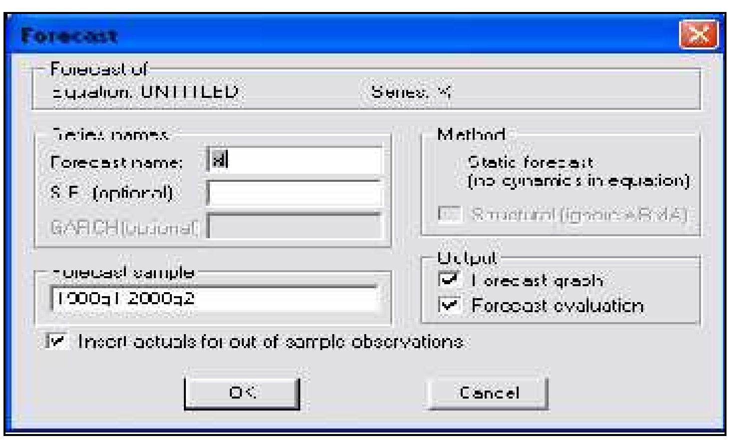
Dependent Variable: X
Method: Least Squares
Date: 05/11/09 Time: 21:23
Sample: 1980:1 2008:2
Included observations: 114

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1647.258	370.3250	-4.448142	0.0000
M	0.982518	0.009436	104.1283	0.0000

R-squared	0.989776	Mean dependent var	29414.66
Adjusted R-squared	0.989685	S.D. dependent var	23069.33
S.E. of regression	2343.009	Akaike info criterion	18.37365
Sum squared resid	6.15E+08	Schwarz criterion	18.42165
Log likelihood	-1045.298	F-statistic	10842.70
Durbin-Watson stat	0.264626	Prob(F-statistic)	0.000000

Ahora, en la ventana de la ecuación anterior vamos al botón de Forecast/Ok., que desplegará una ventana como la siguiente (asegúrese de modificar el Sample de 1980q1 a 2008q2):

Cuadro 12.3.4.4.31: Pronóstico.



En la cual en "Series name/Forecast name/ escribimos Xf/ y el periodo de estimación (Forecast sample) escribimos: 1980q1_2008q2/Ok. Serie que aparecerá en la ventana del Workfile con el nombre especificado: Xf. Una vez hecho esto, debemos volver a predecir con el modelo MCE ya que anteriormente especificamos que sólo predijera a partir del tercer trimestre de 2008 y hasta el cuarto de 2009. Por tanto, siguiendo los pasos descritos anteriormente para predecir a través del MCE, ahora vamos a estimar desde 1980q1 y hasta 2008q2 para realizar la comparación entre ambos métodos. De tal suerte que, una vez que nos encontramos en la ventana de "Model Solution" en "Solution Sample" escribimos 1980q1_2009q4, permaneciendo constantes las demás indicaciones que por default Eviews selecciona/Ok., con lo cual se modificará la serie X_0. Así, seleccionamos la serie original de X, la serie de estimación con el método MCE que por default Eviews denominó como "X_0" y la serie de Xf obtenida con MCO; les damos click derecho y vamos a: Open/As Group/ Ok., lo cual nos desplegará un cuadro comparativo con los siguientes datos:

Cuadro 12.3.4.4.32: Solución del cambio de rango.

Obs	X	X_0	XF
1980:1	5811	5811	5373.81667194
1980:2	6060	6060	6816.15338004
1980:3	6263	6263	7629.67844428
1980:4	6726	6726	8268.31527008
1981:1	8156	8156	8905.96957768
1981:2	8132	8475.513	10152.7851653
1981:3	6989	8813.822	10208.7887023
1981:4	7834	8750.645	10667.6246987
1982:1	7018	9826.429	9270.48382748
1982:2	7709	10130.96	8405.86781717
1982:3	8038	10688.35	6618.66722314
1982:4	8207	10696.36	5333.533426
1983:1	7547	11600.79	4157.45914834
1983:2	8116	11883.02	4942.49118498
1983:3	8270	12594.28	5462.24330935
1983:4	8996	12754.72	5444.55798187
1984:1	9769	13628.96	5814.96734083
1984:2	9502	13934.55	6466.37690314
1984:3	9288	14739.95	7192.45784816
1984:4	9271	15024.82	6995.95420946
1985:1	9245	15920.3	7265.16419448
1985:2	8379	16283.2	7098.13610158
1985:3	9110	17158.11	6994.97169126
1985:4	9124	17543.06	6498.80000353
1986:1	7517	18477.94	6131.33819915
1986:2	7342	18915.47	6403.49573876
1986:3	7025	19852.98	5877.84850522
1986:4	8043	20328.13	5752.08617645
1987:1	8720	21311.49	5566.39023787
1987:2	9549	21831.56	6238.43268224

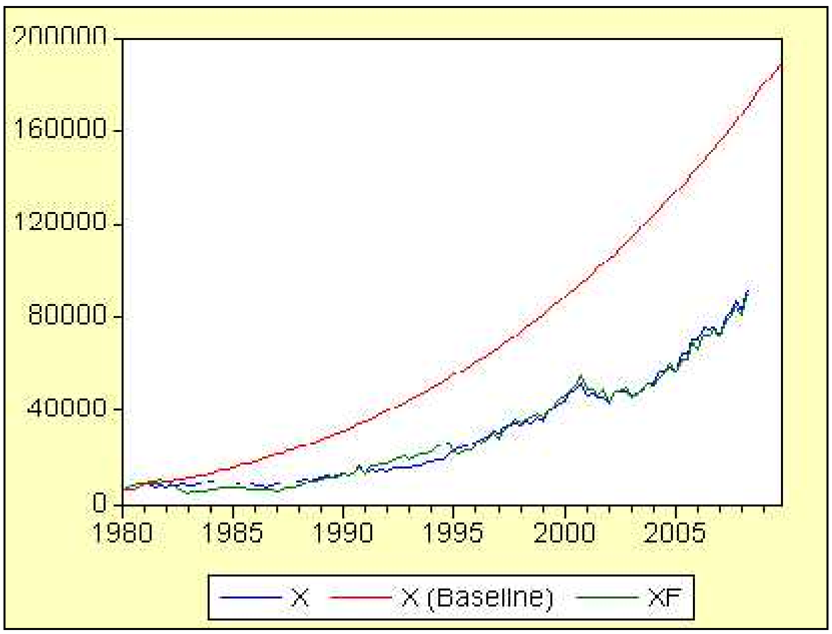
1987:3	9403	22830.84	7069.64307397
1987:4	9697	23393.93	7087.32840145
1988:1	10339	24432.23	7765.26595499
1988:2	10749	25038.83	8878.45906826
1988:3	10343	26101.18	9943.50879005
1988:4	10664	26752.85	10518.2819333
1989:1	11697	27851.37	10857.25071
1989:2	12347	28547.12	11732.6744205
1989:3	11708	29674.88	11882.0171859
1989:4	12351	30416.79	11921.3179136
1990:1	13122	31580.14	13398.0427585
1990:2	12469	32367.32	12233.7586992
1990:3	14204	33563.45	14055.34743
1990:4	16276	34397.72	16136.3209639
1991:1	13391	35630.09	13615.1792793
1991:2	14807	36510.98	16685.5486341
1991:3	14606	37778.83	16788.7130444
1991:4	15283	38707.81	17784.0039744
1992:1	14460	40013.23	17621.8884725
1992:2	15467	40990.22	19344.2428658
1992:3	15484	42333.42	20301.2155863
1992:4	16257	43359.61	20746.2963279
1993:1	15628	44742.04	19269.571483
1993:2	16952	45817.68	20554.7052802
1993:3	16683	47240.01	21293.5589617
1993:4	18489	48366.09	21850.6467775
1994:1	18062	49829.51	22761.4411429
1994:2	19406	51006.5	24764.7957395
1994:3	19461	52511.88	25243.2820997
1994:4	21443	53740.68	26785.8356636

1995:1	23017	55289.12	22298.6750737
1995:2	24056	56570.34	21638.4228476
1995:3	24413	58162.81	22782.0740249
1995:4	25543	59497.31	23574.9662071
1996:1	26626	61134.88	24674.4040657
1996:2	28417	62523.37	26022.4190272
1996:3	29104	64207.08	27797.8294029
1996:4	31169	65650.42	30681.5203009
1997:1	30415	67381.34	28452.1865198
1997:2	32659	68880.32	31604.1048846
1997:3	33334	70659.52	33648.7252454
1997:4	34910	72215.01	36259.2760856
1998:1	33704	74043.58	34689.2120123
1998:2	35476	75656.44	36541.2588071
1998:3	34207	77535.5	36593.3322714
1998:4	36762	79206.63	38998.5368091
1999:1	35424	81137.29	36731.8673367
1999:2	39268	82867.59	39750.1632272
1999:3	40821	84851	41553.0841123
1999:4	43397	86641.42	45212.9643832
2000:1	44326	88678.72	46747.6578015
2000:2	47463	90530.2	48411.0611032
2000:3	49489	92622.57	50929.2552332
2000:4	51598	94536.12	55184.5415294
2001:1	46942	96684.74	49166.617594
2001:2	47806	98661.34	48836.491481
2001:3	45788	100867.4	46605.1926635
2001:4	45630	102908.1	49102.7539114
2002:1	43356	105172.9	44644.0863492
2002:2	48427	107278.7	48743.1522526
2002:3	48090	109603.4	48772.6277984
2002:4	48267	111775.4	49991.9328766
2003:1	46400	114161.3	46212.1853861
2003:2	47911	116400.7	47191.756025
2003:3	49427	118849	48726.4494433
2003:4	51653	121156.8	51677.9340967
2004:1	51683	123668.9	50593.234011
2004:2	57264	126046.3	54157.8100171
2004:3	57071	128623.5	55646.3250803
2004:4	58886	131071.6	60465.5768196
2005:1	56750	133715.2	56625.8957193
2005:2	64774	136235.3	62087.7143571
2005:3	64802	138946.7	62482.6866709
2005:4	70192	141540	69364.2440984
2006:1	69935	144320.6	66546.3819194
2006:2	75941	146988.3	72831.5508034
2006:3	74990	149839.5	72714.6311384
2006:4	76070	152582.9	75255.4231868
2007:1	72625	155506.1	72025.8858847
2007:2	79966	158326.6	78405.3765153
2007:3	82752	161323.2	80639.6228874

2007:4	86569	164222	84335.8563314
2008:1	82609	167293.6	80981.5392187
2008:2	92093	170272.2	90817.5288542
2008:3		173420.2	
2008:4		176479.9	
2009:1		179705.8	
2009:2		182848.1	
2009:3		186153.6	
2009:4		189379.8	

Mismos que si los graficamos se verán así:

Gráfica 12.3.4.4.5: Solución del cambio de rango.



Interpretación: Numérica y gráficamente se observa lo siguiente:

1.- La serie estimada a través de la econometría tradicional (XF) observa un comportamiento similar a los valores de la serie original (X), no así la serie estimada a través del MCE, la cual presenta un comportamiento exponencial debido a que en el proceso de corrección se fueron eliminando las fluctuaciones significativas, dando así origen a una curva alisada con tendencia exponencial, es decir, la estimación se realiza a través del método dinámico, mediante el cual la estimación se realiza a partir del dato anterior estimado y no del valor real.

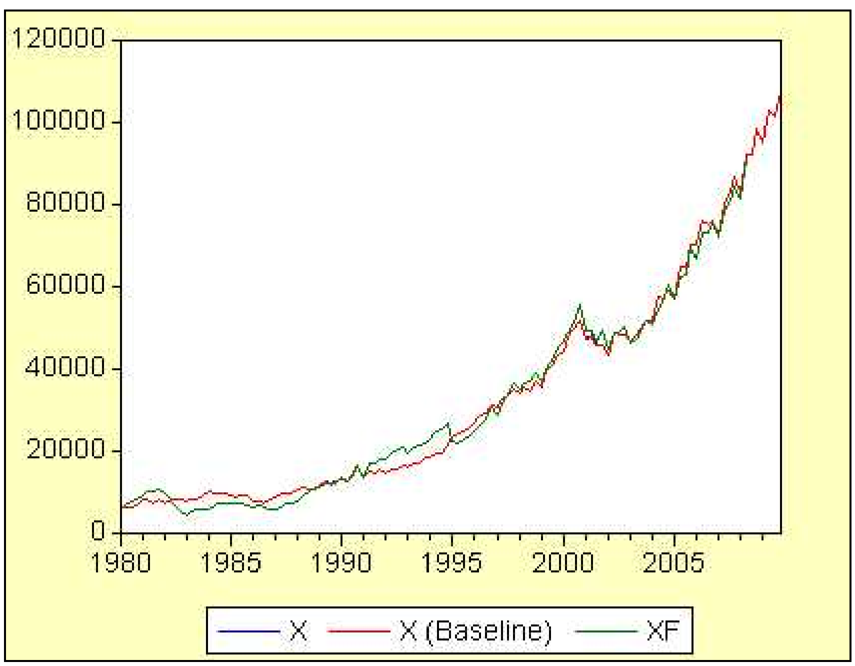
2.- En cada trimestre los valores de XF con respecto a X no difieren sustancialmente en virtud de que XF está estimada a partir de los valores reales de X, en tanto que los de

X_0, como se indicó en el punto anterior, cada valor está estimado con respecto al dato estimado anterior.

3.- El resultado de la estimación para todo el periodo histórico es mejor a través de la econometría tradicional ya que XF sigue un comportamiento similar a X (serie original), en tanto que la estimación para el periodo histórico, con propósitos de predicción, a través del MCE, se aleja de los valores reales debido a que el cálculo de cada una de las observaciones las realiza a partir del valor estimado de la observación estimada anterior, que es la característica del método MCE.

4.- Sin embargo, para fines de predicción, como en el punto anterior, el MCE es adecuado ya que predice, en nuestro caso, para el tercer trimestre de 2008 y hasta el cuarto trimestre de 2009 a partir de los valores reales anteriores de la serie como lo ilustra la ecuación D(X). En contra posición, la predicción para el mismo periodo a través del MCE pero estimado para toda la serie muestra mayores diferencias dado que, como se ha comentado, predice a partir de los valores anteriores estimados de la serie. Gráficamente se observa así:

Gráfica 12.3.4.4.6: Pronóstico.



Comentarios:

1.- Nótese que a diferencia de la gráfica anterior el comportamiento de X_0 (serie estimada a partir del MCE), de color rojo, sigue el mismo comportamiento que la serie original (X), de color azul, es decir, están sobrepuestas, dado que el MCE para predecir, en este caso, utiliza los valores reales y no los estimados como en el caso anterior, utilizado para comparar la bondad en la estimación con respecto a la econometría tradicional.

2.- La serie estimada con econometría tradicional (XF) llega hasta 2008q2 dado que no se puede predecir debido a que no se conocen los valores futuros de la variable independiente, en este caso M. Mientras que la serie X_0 contiene valores futuros porque cada uno de estos se obtienen a partir de los valores rezagados de la serie de acuerdo a la ecuación de predicción.

Para el caso de M, se siguen los mismo pasos descritos para X; se sugiere que el lector los lleva a cabo para afianzar su familiaridad con el tema. Como guía se proporcionan los siguientes resultados:

1. Ecuación de M obtenida con MCO:

Cuadro 12.3.4.4.33: Análisis de Regresión Lineal.

Dependent Variable: M
Method: Least Squares
Date: 05/11/09 Time: 21:31
Sample: 1980:1 2008:2
Included observations: 114

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1982.652	361.0471	5.491394	0.0000
X	1.007387	0.009674	104.1283	0.0000

R-squared	0.989776	Mean dependent var	31614.60
Adjusted R-squared	0.989685	S.D. dependent var	23359.46
S.E. of regression	2372.476	Akaike info criterion	18.39864
Sum squared resid	6.30E+08	Schwarz criterion	18.44665
Log likelihood	-1046.723	F-statistic	10842.70
Durbin-Watson stat	0.266597	Prob(F-statistic)	0.000000

2. Cálculo de M_0 y MF, con el fin de comparar la bondad de ajuste de cada uno de los métodos:

Cuadro 12.3.4.4.34: Solución del cambio de rango.

obs	M	M_0	MF
1980:1	7146	7146	7836.57777715
1980:2	8614	8614	8087.41714646
1980:3	9442	9442	8291.91671261
1980:4	10092	10092	8758.33690535
1981:1	10741	10741	10198.9003516
1981:2	12010	11218.31	10174.723063
1981:3	12067	12107.74	9023.27969302
1981:4	12534	12108.3	9874.52172944

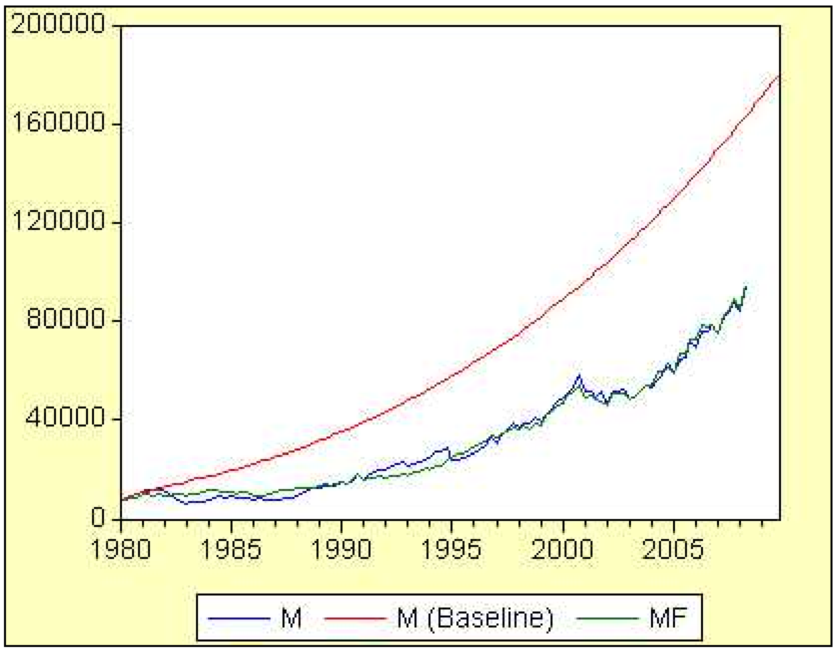
1982:1	11112	13125.29	9052.49391676
1982:2	10232	13514.55	9748.59835128
1982:3	8413	14305.48	10080.0286826
1982:4	7105	14274.49	10250.2770899
1983:1	5908	15225.21	9585.40165317
1983:2	6707	15557.92	10158.6048706
1983:3	7236	16388.73	10313.7424725
1983:4	7218	16487.59	11045.1054529
1984:1	7595	17421.21	11823.8156235
1984:2	8258	17747.83	11554.8432877
1984:3	8997	18635.52	11339.2624643
1984:4	8797	18867.21	11322.1368849
1985:1	9071	19820.61	11295.9448222
1985:2	8901	20181.16	10423.5476583
1985:3	8796	21119.54	11159.9475738
1985:4	8291	21459.43	11174.0509922
1986:1	7917	22445.66	9555.18004241
1986:2	8194	22863.57	9378.88731297
1986:3	7659	23850	9059.54562593
1986:4	7531	24281.94	10085.0656177
1987:1	7342	25307.03	10767.0666339
1987:2	8026	25794.56	11602.1904779
1987:3	8872	26829.74	11455.1119722
1987:4	8890	27346.75	11751.2837577
1988:1	9580	28415.47	12398.026228
1988:2	10713	28979	12811.0549084
1988:3	11797	30064.85	12402.0557761
1988:4	12382	30664.81	12725.4270112
1989:1	12727	31781.6	13766.0578084
1989:2	13618	32424.81	14420.8593749
1989:3	13770	33563.74	13777.1390657
1989:4	13810	34246.73	14424.888923
1990:1	15313	35415.75	15201.5843195
1990:2	14128	36141.18	14543.760592
1990:3	15982	37335.97	16291.577081
1990:4	18100	38103.04	18378.8829975
1991:1	15534	39328.2	15472.5714293
1991:2	18659	40137.99	16899.0314572
1991:3	18764	41391.71	16696.5466651
1991:4	19777	42244.43	17378.5476813
1992:1	19612	43529.49	16549.4681594
1992:2	21365	44425.67	17563.906894
1992:3	22339	45741.61	17581.0324734
1992:4	22792	46681.83	18359.742644
1993:1	21289	48030.46	17726.0962051
1993:2	22597	49015.14	19059.8766266
1993:3	23349	50396.72	18788.8895168
1993:4	23916	51426.53	20608.2304846
1994:1	24843	52842.4	20178.0762248
1994:2	26882	53917.78	21532.0043869
1994:3	27369	55368.55	21587.4106732
1994:4	28939	56490.19	23584.0517575

1995:1	24372	57976.99	25169.6789354
1995:2	23700	59145.41	26216.3540548
1995:3	24864	60669.02	26575.9912228
1995:4	25671	61884.91	27714.3385615
1996:1	26790	63446.38	28805.3387099
1996:2	28162	64710.31	30609.5688723
1996:3	29969	66310.52	31301.6437587
1996:4	32904	67623.21	33381.8979661
1997:1	30635	69263.15	32622.328149
1997:2	33843	70625.25	34882.9046339
1997:3	35924	72305.87	35562.890876
1997:4	38581	73718.1	37150.532828
1998:1	36983	75440.4	35935.6240754
1998:2	38868	76903.46	37720.7138843
1998:3	38921	78668.42	36442.3397491
1998:4	41369	80183.07	39016.2135989
1999:1	39062	81991.7	37668.329759
1999:2	42134	83558.68	41540.7254845
1999:3	43969	85412	43105.1975348
1999:4	47694	87032.1	45700.2265121
2000:1	49256	88931.14	46636.0890587
2000:2	50949	90605.16	49796.2621572
2000:3	53512	92550.97	51837.2282706
2000:4	57843	94279.72	53961.8075071
2001:1	51718	96273.35	49271.413517
2001:2	51382	98057.68	50141.7959069
2001:3	49111	100100.2	48108.8888898
2001:4	51653	101941	47949.7217398
2002:1	47115	104033.5	45658.9236441
2002:2	51287	105931.6	50767.3832497
2002:3	51317	108075.2	50427.8938221
2002:4	52558	110031.5	50606.2013256
2003:1	48711	112227.4	48725.4097493
2003:2	49708	114242.9	50247.5715446
2003:3	51270	116492	51774.770275
2003:4	54274	118567.6	54017.2137935
2004:1	53170	120871.2	54047.4354042
2004:2	56798	123008	59669.6623927
2004:3	58313	125367.2	59475.2366968
2004:4	63218	127566	61303.6441479
2005:1	59310	129982.2	59151.8654617
2005:2	64869	132244	67235.1389532
2005:3	65271	134718.2	67263.3457899
2005:4	72275	137044.2	72693.1618565
2006:1	69407	139577.7	72434.263391
2006:2	75804	141968.9	78484.6298653
2006:3	75685	144562.9	77526.6048042
2006:4	78271	147020.4	78614.5827916
2007:1	74984	149676.1	75144.1344892
2007:2	81477	152201.1	82539.3626424
2007:3	83751	154919.8	85345.942895
2007:4	87513	157513.3	89191.1391708

2008:1	84099	160296.4	85201.8865504
2008:2	94110	162959.6	94755.9450988
2008:3		165808.3	
2008:4		168542.5	
2009:1		171458.2	
2009:2		174264.4	
2009:3		177248.5	
2009:4		180128.1	

Mismos que si los graficamos, obtenemos:

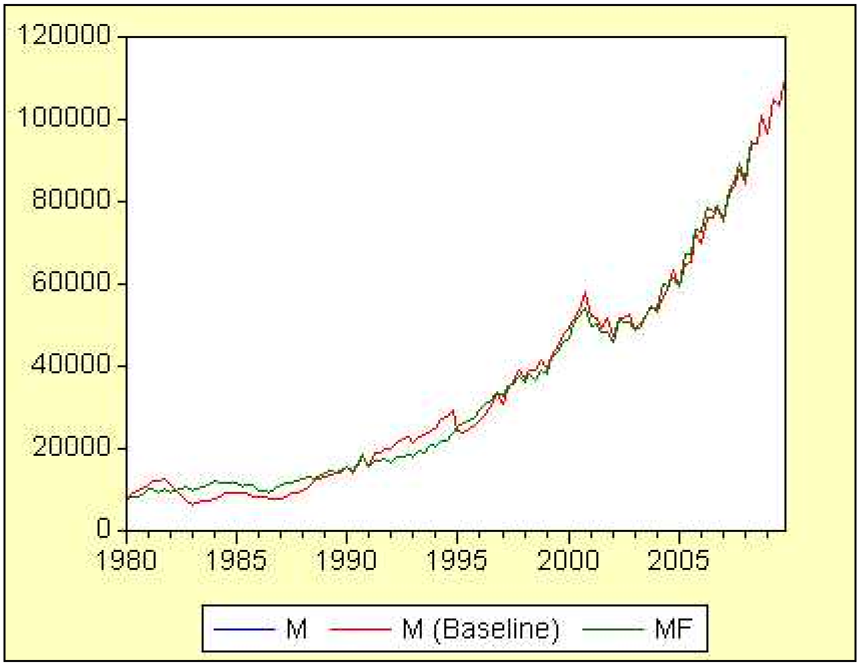
Gráfica 12.3.4.4.7: Solución del cambio de rango.



3. Estos resultados se deben interpretar en forma similar a X, destacando que la econometría tradicional sirve principalmente para análisis de estructura y que el MCE para análisis de predicción.

4. Estos resultados gráficamente se observan así y revelan un comportamiento similar al de X, es decir, M_0 está sobrepuesta en M por lo cual no se observa; igualmente la estimación con econometría tradicional (curva color verde) llega hasta 2008q2, por las mismas razones antes señaladas.

Gráfica 12.3.4.4.8: Pronóstico.



Conclusiones:

1. Para fines de predicción el MCE es una mejor opción en tanto que la econometría tradicional se ve limitada dado que no se conocen los valores futuros de la o las variables independientes.
2. Para análisis de estructura y de evaluación de políticas públicas o privadas, la econometría tradicional resulta la mejor alternativa ya que utiliza los valores reales de la serie.

XII.4.-PRONÓSTICO.

XII.4.1.- Predicción univariante: de los valores de la serie de tiempo Y_t.

Una vez que se ha identificado la existencia o tipo de variaciones (tendencia, estacional, cíclica y residual) que puede presentar Y_t, y que dichas oscilaciones se hayan eliminado con cualesquiera de los métodos antes señalados, **ya sean los que hemos usado para eliminar la estacionalidad o con los aplicados para lograr la estacionariedad;** derivado de lo anterior ahora se está en condiciones de hacer predicciones de los valores de Y_t.

En este contexto, las predicciones pueden ser de dos tipos: *condicionales o incondicionales*.

Las condicionales tienen como fundamento la relación causal que existe entre la variable explicada Y_t y su explicativa X_t, de manera que Y_t se puede predecir sí y sólo sí se conocen los valores futuros de X_t.

Las incondicionales se realizan usando métodos autoproyectivos que, como su nombre lo indica, para predecir los valores de Y_t sólo se necesita conocer sus valores pasados y el actual, por lo que también se conocen como autorregresivos, AR, con los que se hallarán sus valores futuros; en otras palabras, no es necesario conocer los valores de X_t. Estos métodos pueden basarse en dos enfoques alternativos:

3. El determinista o clásico; y
4. El estocástico o moderno basado en Box and Jenkins.

Se recomienda el determinista cuando la muestra es pequeña; el estocástico cuando la muestra es grande.

Derivado de las recomendaciones anteriores es que decimos que según sea la predicción es que se seguirán ciertos métodos, que son más apropiados para predecir. Ejemplo:

3. Para el largo plazo se sugiere el análisis de tendencia;
4. Para el corto y mediano plazo los métodos econométricos.

Al respecto, para la proyección a corto plazo se deben de tener en cuenta las variaciones estacionales, E_t. Para el mediano plazo, las oscilaciones cíclicas.

XII.4.2.- Consideraciones.

Durante la exposición de la teoría econométrica a lo largo de los capítulos que integran este libro, hemos proporcionado el marco teórico necesario con los conceptos básicos para poder modelar relaciones entre variables económicas usando la metodología econométrica, tanto la tradicional como la moderna, de manera que el lector ya está en condiciones de plantear una teoría económica, formularla matemáticamente, verificarla estadísticamente y así proyectar las variables que comprende el modelo. Agréguese a lo anterior que en los ejercicios que haremos enseguida para constatar y afianzar el uso de los métodos econométricos, también se le está orientando sobre cómo disponer de las bases de datos de las variables macroeconómicas así como acerca de las operaciones que se deben realizar al usar el software EViews para obtener los resultados, cuyo análisis e interpretación es casi seguro que serán correctas puesto que se cuenta con el marco teórico suficiente para hacerlo.

XII.4.3.- EJERCICIOS CON EVIEWS.

A. Aspectos básicos con aplicaciones a la economía mexicana (Rojas: 2009).

En esencia este es un resumen bastante ilustrativo y práctico de los conceptos y métodos que constituyen la econometría de series temporales, cuya ilustración se hace aplicándolos desde sus orígenes con el análisis univariante propuesto en los modelos ARIMA y ARMA de Box & Jenkins, pasando por su profundización y extensión a más variables que hicieron Granger y Eagle con su teoría y metodología de la cointegración, hasta llegar al análisis de predicción de los valores de las series temporales en el largo plazo.

Por lo anterior es que el ejercicio se inicia con el engarzamiento de los conceptos básicos de dicha metodología aplicados desde el principio a los modelos ARIMA. Así, como vimos en el marco teórico descrito en los capítulos 10 y 11, el desarrollo de una aplicación de tipo ARIMA a través de la metodología Box & Jenkins debe partir necesariamente de un proceso estocástico estacionario, es decir, $\hat{\circ}_k \sim N(0, 1/n)$. Por lo que el primer paso a realizar es el análisis de la estacionariedad de dicho proceso, prestando especial atención al comportamiento constante en media y varianza (Pulido y Pérez, 2000: 167), dado que la metodología Box & Jenkins precisa que las series sean estacionarias, por lo que es importante comprobar esta situación antes de trabajar con ellas en el análisis de predicción, tema central de esta sección, de gran uso al hacer economía aplicada con econometría empírica y epilogo de toda la metodología comprometida en este libro.

Pero, reiteramos, ¿Porqué las series de tiempo estacionarias son tan importantes? De acuerdo a Gujarati (2007: 772-773) porque, si una serie de tiempo es no estacionaria, se puede estudiar su comportamiento sólo durante el periodo bajo consideración. Por tanto, cada conjunto de datos perteneciente a la serie de tiempo corresponderá a un episodio particular. Como consecuencia, no puede generalizarse para otros periodos. Así, pues, para propósitos de pronóstico (predicción a largo plazo), tales series de tiempo (no estacionarias) tendrán un valor práctico insignificante.Derivado de lo anterior es que se debe procurar que éstas sean estacionarias, es decir, que observen estabilidad estadística en el tiempo en términos de media, varianza y covarianza, minímamente, ya que para hacer planeación a largo plazo, como es el Plan Nacional de Desarrollo de México, en el cual se plasma la visión del México que todos queremos cada seis años; en él se expresan las relacionan y por consiguiente la evolución de las variables que se estima producirán el crecimiento y desarrollo sostenido y sustentable anhelado históricamente por los mexicanos.

Identificación de la estacionariedad de la serie.

Existen dos formas de verificar la estacionariedad del proceso o de las variables aleatorias, a saber:

1.-El método informal (gráfico); y
2.-El método formal (las pruebas de raíz unitaria).

El modelo ARIMA

Cabe destacar que para ilustrar la aplicación de la metodología ARIMA con Eviews, como ya lo hicimos en el Capítulo 11, vamos a trabajar con el precio de las acciones de Consorcio ARA, la cual cotiza en la Bolsa Mexicana de Valores, para el periodo que comprende desde el 2 de enero de 2008 hasta el 19 de febrero de 2009, es decir, una muestra de 285 datos.

1.-Método informal.
Creación de un nuevo archivo:

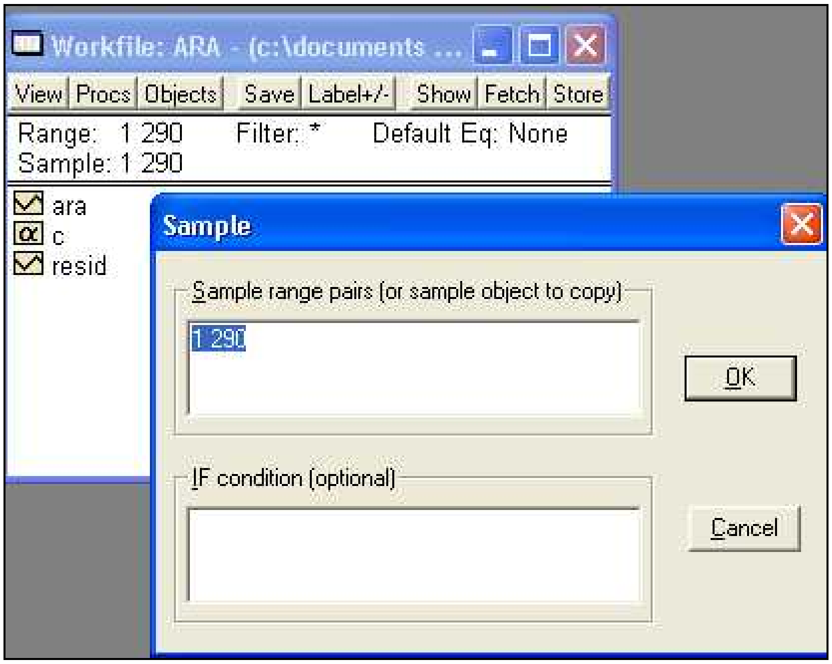
Para generar un nuevo archivo de trabajo (Workfile) en Eviews , una vez que se ha abierto la aplicación, en el menú principal seleccionamos: File/new/workfile/ok, con lo que se desplegará la ventana de información ya conocida. Como se trata del precio de acciones y dado que este mercado labora de lunes a viernes y se mantiene cerrado los sábados y domingos así como días festivos, lo lógico sería que seleccionáramos la frecuencia de datos diarios con semanas de 5 días (Daily [5 day weeks]). Desgraciadamente Eviews no reconoce los días festivos del calendario mexicano, por lo que nos marcará un tamaño de muestra más grande que la que poseemos. Sin embargo, en caso de que el lector desee realizar la prueba, en Eviews una vez que selecciono la frecuencia de los datos, en el rango debe escribir el periodo en el siguiente orden: mes:día:año, con el siguiente formato: mm:dd:aaaa. De tal suerte que, si el periodo de inicio es el 2 de enero de 2008, escribiremos en “Start date”: 01:02:2008 y si el periodo final de la muestra es el 19 de febrero de 2009 escribiremos en “End date”: 02:19:2009.

Para nuestros fines vamos a seleccionar frecuencia de datos irregulares (Undated or irregular) y dado que el tamaño de nuestro proceso estocástico es de 285 variables aleatorias en “Start date” vamos a escribir 1 y en “End date” 290 porque vamos a predecir el precio de las acciones de ARA para cinco días más. Así, 1 se referirá a la variable aleatoria del 2 de enero de 2008, 285 a la del 19 de febrero de 2009 y de la variable 286 a la 290 se referirá a la estimación de dicha variable, por lo que la variable 286 se refiere al precio del 20 de febrero, 287 a la del 23, 288 a la del 24, 289 a la del 25 y 290 a la del 26 de febrero de 2009.

Una vez generado el Workfile vamos al menú principal y ahí seleccionamos: Quick/Empty Group (Edit Series)/Ok. Lo cual nos desplegará una ventana de datos y en la cual capturaremos todas las variables aleatorias que componen el proceso estocástico. Mismo que denominaremos como “ARA”. Dicho proceso aparecerá en la ventana del Workfile.

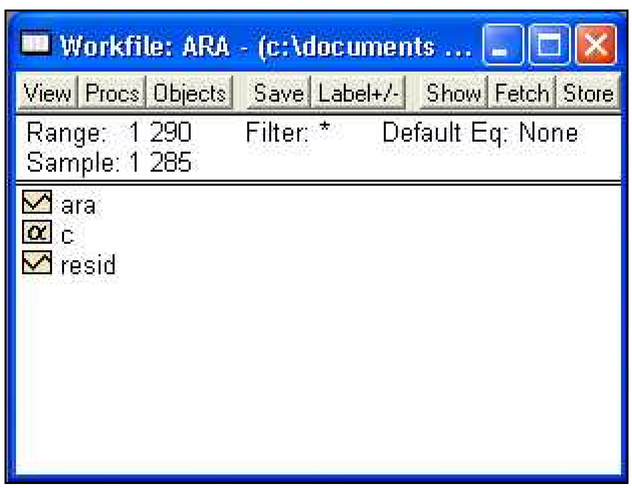
Una vez cargada la serie “ARA”, para empezar a trabajar cambiamos el periodo muestral (Sample) en la ventana del Workfile dando doble click sobre el mismo, es decir, sobre “Sample”. Este procedimiento desplegará una ventana denominada “Sample” que se observa de la siguiente forma:

Cuadro 12.4.3.1.: Modificación de la muestra.



En la ventana del Sample cambiamos el 290 por el 285. En la ventana del Workfile se observa ya el cambio realizado. Mientras que el Rango (Range) se mantiene inalterado. Lo cual presenta una estructura como esta:

Cuadro 12.4.3.2.: Cambio realizado en el simple.



e) Gráfica lineal y diagrama de dispersión.

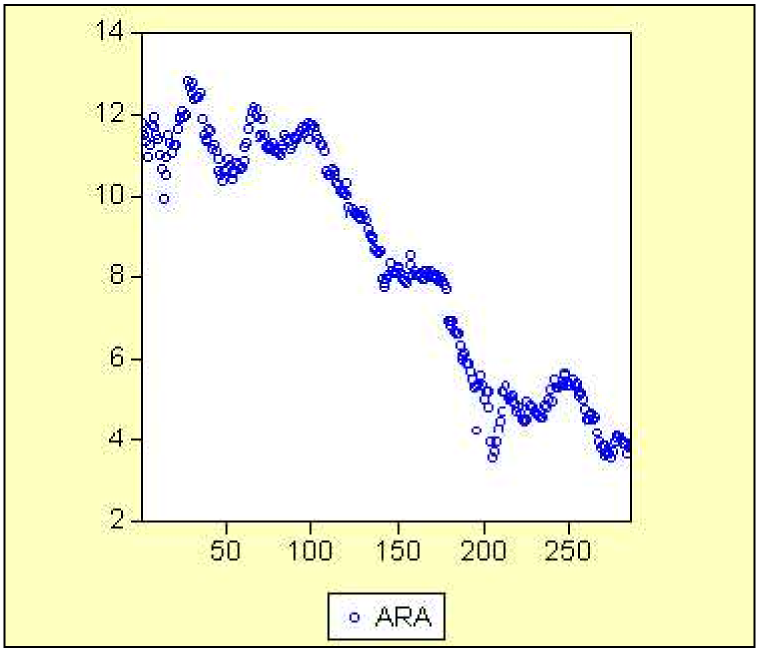
Hecho esto, abrimos la serie “ARA” dando doble click sobre ella en la ventana del workfile. Una vez abierta la ventana de la serie vamos a: View/Graph/Line/Ok. Lo que nos desplegará una gráfica como la siguiente:

Gráfica 12.4.3.1.: Comportamiento gráfico de la variable “ARA”

Teóricamente, si el proceso estocástico muestra decrecimiento o crecimiento significa que la serie no presenta un valor medio constante en todo el periodo muestral, es decir, no oscila en torno al mismo valor. En otros términos, una serie es estacionaria cuando su media y varianza permanecen constantes en el tiempo. Cuando se comprueba que una serie no presenta media y varianza constantes, entonces se dice que no es estacionaria. Por tanto, podemos suponer, a priori, que, probablemente, no es estacionaria, luego, presentará al menos una raíz unitaria (Pulido y López, 1999: 269, 273).

Comentario: Como puede observarse, la gráfica del proceso estocástico ARA presenta un claro comportamiento descendente o decreciente por lo que, podemos suponer que no es constante en media, luego entonces, podemos suponer que el proceso no es estacionario. Además, existen variables dentro del proceso que muestran gran dispersión, por lo que tampoco es constante en varianza (ver diagrama de dispersión en el cuadro siguiente). Derivado de lo cual, quizá sea necesario y/o preciso trabajar en diferencias. Además, tomando logaritmos se reduce la dispersión de la serie, es decir, la transformación en logaritmos la convierte en estacionaria en varianza. Sin embargo, realicemos otras pruebas gráficas y formales para determinarlo con precisión.

Gráfica 12.4.3.2.: Dispersión de la variable "ARA"

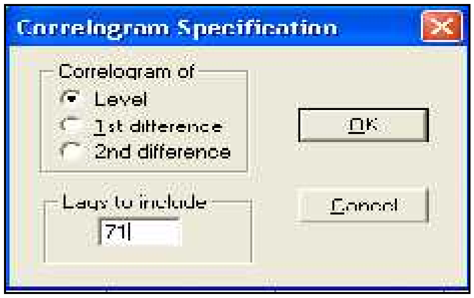


Nota: Para obtener el diagrama de dispersión de el proceso estocástico ARA vamos al menú principal y ahí seleccionamos: Quick/Graph/Scatter/, donde se desplegará una pantalla y en la cual escribiremos la serie a graficar, en nuestro caso ARA, y le damos Ok.

f) Correlograma.

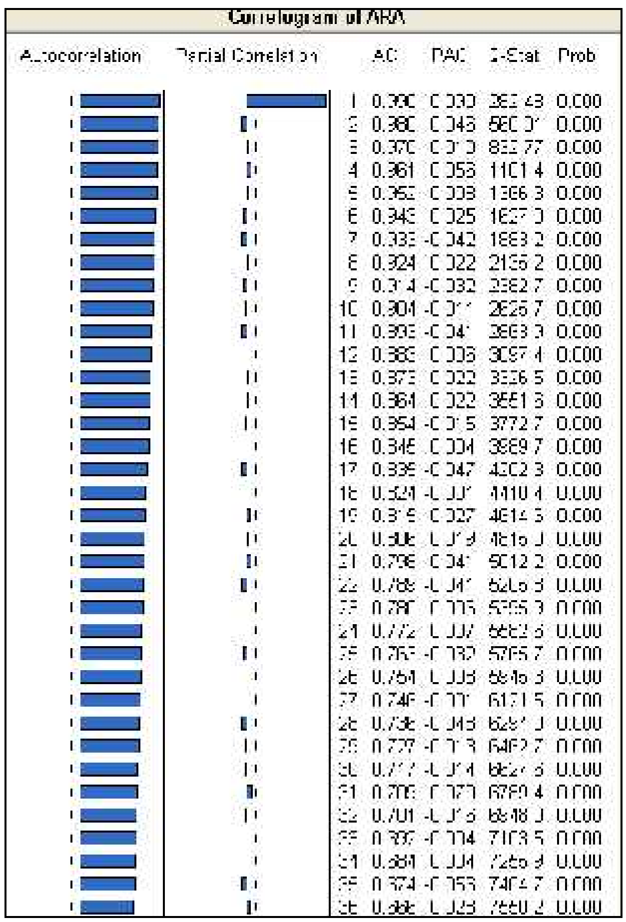
Otro procedimiento gráfico para comprobar la posible existencia de una raíz unitaria en el proceso estocástico consiste en inspeccionar el correlograma del mismo. Así, en la ventana de la serie vamos a: View/Correlogram…/, en donde se desplegará una ventana que nos solicita la especificación del correlograma a realizar, las opciones son: en niveles (Level), es decir, los valores de el proceso original, en primeras diferencias y en segundas diferencias. Asimismo, nos solicita en número de rezagos a incluir (Lags to include), mismo que por default Eviews da 36. La ventana es como la siguiente:

Cuadro 12.4.3.3.: Especificación de correlograma.



Sin embargo, siguiendo el criterio y/o sugerencia de Gujarati (2007: 784-785) en cuanto a el número de rezagos a utilizar y/o a la longitud del mismo, utilizaremos un cuarto de la longitud de la serie de tiempo, es decir, 71 variables aleatorias. Por lo que, para nuestro caso seleccionaremos el correlograma de niveles (Level) y en "Lags to include" escribiremos 71/Ok. Procedimiento que nos desplegará el correlograma siguiente:

Cuadro 12.4.3.4.: Correlograma de "ARA".



Correlogram of ARA

	AC	PAC	Q-Stat	Prob
1	0.990	0.990	283.43	0.000
2	0.980	0.043	[illegible]	0.000
3	0.970	[illegible]	833.77	0.000
4	0.961	0.053	1101.4	0.000
5	0.952	0.008	1366.3	0.000
6	0.943	0.025	1627.0	0.000
7	0.933	-0.042	1883.2	0.000
8	0.924	0.022	2135.2	0.000
9	0.914	-0.032	2382.7	0.000
10	0.904	[illegible]	2625.7	0.000
11	0.893	[illegible]	[illegible]	0.000
12	0.883	0.006	3097.4	0.000
13	0.873	0.022	3326.5	0.000
14	0.864	0.022	3551.3	0.000
15	0.854	-0.015	3772.7	0.000
16	0.845	0.004	3989.7	0.000
17	0.835	-0.047	4202.3	0.000
18	0.824	[illegible]	4410.4	0.000
19	0.815	0.027	4614.3	0.000
20	0.806	[illegible]	[illegible]	0.000
21	0.796	[illegible]	5012.2	0.000
22	0.788	[illegible]	[illegible]	0.000
23	0.780	[illegible]	[illegible]	0.000
24	0.772	0.007	[illegible]	0.000
25	0.763	-0.032	5765.7	0.000
26	0.754	0.003	[illegible]	0.000
27	0.746	[illegible]	6121.5	0.000
28	0.738	-0.043	[illegible]	0.000
29	0.727	[illegible]	[illegible]	0.000
30	[illegible]	[illegible]	[illegible]	0.000
31	0.706	[illegible]	[illegible]	0.000
32	[illegible]	[illegible]	[illegible]	0.000
33	[illegible]	-0.004	7103.5	0.000
34	0.681	0.004	[illegible]	0.000
35	0.674	-0.053	[illegible]	0.000
36	[illegible]	0.023	[illegible]	0.000

En términos teóricos, si la función de autocorrelación muestral (primera columna de la izquierda del correlograma) decrece lentamente es muestra o indicio de la no estacionariedad de la serie, es decir, presenta una raíz unitaria. Asimismo, si la autocorrelación parcial (segunda columna de la izquierda del correlograma) muestra al menos un valor significativo en el retardo uno, con un coeficiente de autocorrelación parcial (PAC) cercano a la unidad ($\rho \sim 1$) es indicativo de la no estacionariedad de la serie (Pulido y López, 1999: 274).

Ahora bien, un correlograma que desciende rápidamente o es cuasialeatorio corresponde a variables estacionarias.

Asimismo, el estadístico “Q” también es una prueba de utilidad para determinar la no estacionariedad de la serie, pues si la probabilidad asociada a los rezagos es menor que 0.05 se dice que la serie es no estacionaria y, si la probabilidad asociada es mayor a 0.05 se concluye que la serie es estacionaria.

Sencillamente, cuando en el correlograma la función de autocorrelación desciende lentamente, así como si los valores de la autocorrelación muestral (AC) de los primeros rezagos son cercanos a 1, los primeros de la autocorrelación parcial (ACP) son cercanos a 1 y la probabilidad asociada a la estadística “Q” para cada rezago es menor a 0.05, no cabe duda de que se trata de una serie de tiempo no estacionaria.

Comentario: Dado que como puede observarse en el correlograma de la serie ARA la función de autocorrelación desciende lentamente y la función de autocorrelación parcial presenta un valor significativo en el primer coeficiente, cercano a la unidad, debemos proceder a la transformación de la serie en logaritmos y posteriormente a diferenciarla. Asimismo, como todas las probabilidades asociadas al estadístico “Q” son menores a 0.05 se concluye que la serie es no estacionaria.

3. Método formal.

El método formal para analizar si una serie es estacionaria es el Test de raíces unitarias. Eviews incluye dos pruebas para probar la existencia de raíz unitaria, se trata de la prueba o Test de Dickey-Fuller (DF), una variante del mismo la Dickey-Fuller Aumentada (ADF) y el Test de Phillips-Perron (PP).

Para ambas pruebas se plantea la siguiente prueba de hipótesis:

Ho: $\rho = 1$; existe raíz unitaria, por tanto, la serie es no estacionaria.

Ha: $\rho < 1$; no existe raíz unitaria, por tanto, la serie es estacionaria.

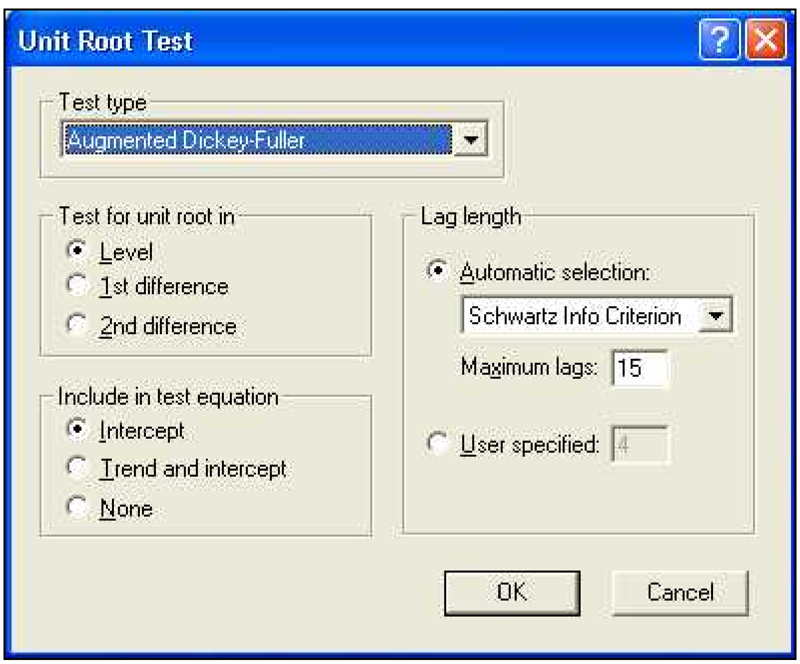
El criterio de aceptación de la Ho es que si el valor de t del estadístico de DF, ADF y PP es menor que los valores de t críticos de Mackinnon (a cualquier nivel de significación al que estemos trabajando), ambos en términos absolutos, se acepta la Ho. Si el valor de los estadísticos es mayor que los valores críticos de Mackinnon, ambos en términos absolutos, se acepta la Ha.

f) Test DF y ADF.

En Eviews para realizar la prueba DF y ADF una vez estando en la ventana de la serie vamos a: View/ Unit Root Test…/, en donde nos aparecerá una ventana como la siguiente. Cabe destacar que el Test de DF o simple es solamente válido si las series son un proceso autorregresivo de orden 1, AR(1).

El Test de ADF se utiliza cuando se tienen retardos o rezagos de orden superior al 1, por tanto, la serie sigue un proceso AR(ρ).

Cuadro 12.4.3.5.: Raíz Unitaria Prueba de Dickey-Fuller Aumentada.



En primer lugar, hay que especificar el número de retardos de los términos de las primeras diferencias de la serie para incluir en la regresión del Test. En Eviews tenemos dos opciones para seleccionar el número de retardos (Lag length): De forma automática (Automatic selection) y especificado por el usuario (User specified). Si, por ejemplo, seleccionamos el especificado por el usuario e incluimos cero (0) retardos o rezagos, Eviews nos proporciona el Test de DF y si indicamos un número mayor que cero, generará el Test de ADF.

En segundo lugar, el programa pregunta acerca de la inclusión de otras variables exógenas en la regresión auxiliar del Test. Podemos incluir una constante (Intercept), una constante y un término de tendencia lineal (Trend and intercept) o, simplemente, no incluir nada adicional en la regresión del Test (None). ¿Qué variables incluir entonces? Podemos aplicar las siguientes normas de carácter general:

j) Si la serie original presenta tendencia (creciente o decreciente), se deberán incluir ambos términos en la regresión, es decir, considerar como regresores al término independiente y al término de tendencia lineal;
ii) Si la serie original no parece presentar tendencia y tiene un valor medio distinto de cero debemos incluir un término constante en la regresión; y
iii) Si la serie original parece fluctuar en torno al valor medio cero, no se considera necesario incluir ningún regresor adicional en la regresión, es decir, no incluimos ni constante ni término de tendencia (Pulido y López, 1999: 271).

Entonces, una vez estando en la ventana de la serie, como se comento previamente, vamos a: View/Unit Root Test…/, con lo que aparecerá una ventana de dialogo como la anterior, en la cual seleccionaremos el Test de ADF teniendo siempre presente las consideraciones anteriores. Para aplicar el Test de DF:

- Seleccionamos "Augmented Dickey-Fuller" en la especificación del Test (Test Type);
- Después, indicamos si el Test se va a realizar sobre la serie en niveles u original (Level), en primeras diferencias o en segundas diferencias de la serie original en "Test for unit root in";
- Especificamos si queremos incluir una constante (intercept), una constante más un término de tendencia lineal (Trend and intercept) o ninguno de los dos casos anteriores (None); y
- Finalmente, debemos especificar el orden de la correlación serial a considerar en las series. Para el Test de ADF, se trata de especificar el número de retardos de los términos de primeras diferencias de la serie a añadir en la regresión del Test (Pulido y López, 1999: 272).

Algunas consideraciones generales sobre la selección de los datos es que es muy importante y hay que considerar lo siguiente:

j) Si el Test acepta la Ho en los datos en niveles, pero rechaza la Ho en primeras diferencias, entonces la serie contiene una raíz unitaria y, por tanto, es integrada de orden 1, I(1); y

ii) Si el Test acepta la Ho en niveles y primeras diferencias, pero se rechaza al realizar el Test en segundas diferencias, entonces la serie contiene dos raíces unitarias y es integrada de orden 2, I(2).

Para nuestro caso, vamos a realizar los dos Test, es decir, el DF y el ADF.

- **Test DF.**

Dado que la serie original de ARA presenta tendencia decreciente vamos a incluir en el Test el intercepto y el término de tendencia lineal; luego, vamos a seleccionar los rezagos de forma manual o especificados por el usuario, en donde indicaremos que son cero para trabajar con el Test de DF, y; finalmente, vamos a realizar el Test primero en niveles, luego en primeras diferencias para aceptar o rechazar la Ho. Así, siguiendo los pasos preestablecidos, el resultado para el Test de DF en niveles es:

Cuadro 12.4.3.6.: Análisis de Regresión de la Prueba de Dickey-Fuller Aumentada.

Null Hypothesis: ARA has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-2.438438**	**0.3589**
Test critical values:	**1% level**	**-3.990701**	
	5% level	**-3.425728**	
	10% level	**-3.136027**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ARA)
Method: Least Squares
Date: 03/01/09 Time: 23:50
Sample(adjusted): 2 285
Included observations: 284 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ARA(-1)	-0.036564	0.014995	-2.438438	0.0154
C	0.466007	0.199219	2.339171	0.0200
@TREND(1)	-0.001323	0.000536	-2.470401	0.0141

R-squared	0.021590	Mean dependent var	-0.027254
Adjusted R-squared	0.014627	S.D. dependent var	0.241696
S.E. of regression	0.239922	Akaike info criterion	-0.006496
Sum squared resid	16.17512	Schwarz criterion	0.032050
Log likelihood	3.922429	F-statistic	3.100400
Durbin-Watson stat	1.890580	Prob(F-statistic)	0.046575

Comentario: Como el valor de de t de DF, 2.43, es menor que el valor t crítico de MacKinnon al 5%, en términos absolutos, de significación estadística, 3.42, se acepta la Ho. Luego, entonces, existe raíz unitaria, por tanto, la serie es no estacionaria. Asimismo, al ser la probabilidad asociada al estadístico DF mayor que 0.05 se acepta la Ho y se rechaza la Ha.

Ahora, realicemos el Test de DF pero en primeras diferencias con los demás criterios constantes, así obtenemos:

Cuadro 12.4.3.7.: Análisis de Regresión de la Prueba de Dickey-Fuller Aumentada.

Null Hypothesis: D(ARA) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-16.12400**	**0.0000**
Test critical values:	**1% level**	**-3.990817**	
	5% level	**-3.425784**	
	10% level	**-3.136061**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ARA,2)
Method: Least Squares
Date: 03/01/09 Time: 23:57
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ARA(-1))	-0.962111	0.059669	-16.12400	0.0000
C	-0.017313	0.029055	-0.595879	0.5517
@TREND(1)	-6.78E-05	0.000176	-0.384186	0.7011

R-squared	0.481487	Mean dependent var	-0.000353
Adjusted R-squared	0.477784	S.D. dependent var	0.335380
S.E. of regression	0.242361	Akaike info criterion	0.013768
Sum squared resid	16.44689	Schwarz criterion	0.052412
Log likelihood	1.051834	F-statistic	130.0031
Durbin-Watson stat	1.984524	Prob(F-statistic)	0.000000

Comentario: Dado que en términos absolutos el valor del estadístico DF, 16.12, es mayor que el valor crítico de MacKinnon al 5% de significación igual a 3.42, se rechaza la Ho y se acepta la Ha, por tanto, no existe raíz unitaria, es decir, la serie es estacionaria. Asimismo, al ser la probabilidad asociada al estadístico DF menor a 0.05 se acepta la Ha y se rechaza la Ho.

Conclusión preliminar: En virtud de que el Test de DF en primeras diferencias rechaza la Ho y acepta la Ha y dado que en niveles se acepta la Ho y se rechaza la Ha concluimos que la serie contiene una raíz unitaria y, por tanto, es integrada de orden 1, I(1).

- **Test ADF.**

Para elaborar este Test vamos a modificar únicamente el número de rezagos, así, en lugar de incluir cero como en el Test anterior de DF, vamos a incluir 1. Los demás términos los mantendremos constantes, de tal suerte que como la serie original de ARA presenta tendencia decreciente vamos a incluir en el Test el intercepto y el término de tendencia lineal; luego, vamos a seleccionar los rezagos de forma manual o especificados por el usuario, en donde indicaremos que es uno para trabajar con el Test de ADF, y; finalmente, vamos a realizar el Test primero en niveles y luego en primeras diferencias para aceptar o rechazar la Ho. Así, siguiendo los pasos descritos, el resultado para el Test de ADF en niveles es:

Cuadro 12.4.3.8.: Análisis de Regresión de la Prueba de Dickey-Fuller Aumentada.

Null Hypothesis: ARA has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-2.445320**	**0.3554**
Test critical values:	**1% level**	**-3.990817**	
	5% level	**-3.425784**	
	10% level	**-3.136061**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ARA)
Method: Least Squares
Date: 03/02/09 Time: 01:21
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ARA(-1)	-0.037107	0.015175	-2.445320	0.0151
D(ARA(-1))	0.053875	0.059506	0.905367	0.3661
C	0.471624	0.202012	2.334639	0.0203
@TREND(1)	-0.001325	0.000543	-2.439810	0.0153

R-squared	0.022963	Mean dependent var	-0.028057
Adjusted R-squared	0.012457	S.D. dependent var	0.241745
S.E. of regression	0.240234	Akaike info criterion	-0.000371
Sum squared resid	16.10179	Schwarz criterion	0.051155
Log likelihood	4.052452	F-statistic	2.185769
Durbin-Watson stat	1.983966	Prob(F-statistic)	0.089938

Comentario: Dado que el valor del estadístico de ADF es menor en términos absolutos al valor crítico de MacKinnon, a saber 2.44 < 3.42 al 5% de significación estadística, y en virtud de que la probabilidad asociada al estadístico ADF es mayor a 0.05 se acepta la Ho y por tanto se dice que existe raíz unitaria, es decir, la serie es no estacionaria.

El resultado del Test de ADF con primeras diferencias manteniendo constante los demás términos es:

Cuadro 12.4.3.9.: Análisis de Regresión de la Prueba de Dickey-Fuller Aumentada.

Null Hypothesis: D(ARA) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-11.88449**	**0.0000**
Test critical values:	**1% level**	**-3.990935**	
	5% level	**-3.425841**	
	10% level	**-3.136094**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(ARA,2)
Method: Least Squares
Date: 03/02/09 Time: 01:36
Sample(adjusted): 4 285
Included observations: 282 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ARA(-1))	-0.986810	0.083033	-11.88449	0.0000
D(ARA(-1),2)	0.029472	0.059800	0.492843	0.6225
C	-0.013561	0.029296	-0.462898	0.6438
@TREND(1)	-9.19E-05	0.000178	-0.517239	0.6054

R-squared	0.480514	Mean dependent var	0.001383
Adjusted R-squared	0.474908	S.D. dependent var	0.334700
S.E. of regression	0.242534	Akaike info criterion	0.018737
Sum squared resid	16.35279	Schwarz criterion	0.070395
Log likelihood	1.358073	F-statistic	85.71467
Durbin-Watson stat	2.008458	Prob(F-statistic)	0.000000

Comentario: Como puede observarse, al ser el valor del Test ADF mayor que el valor crítico de MacKinnon (11.88 > 3.42) en términos absolutos al 5% de significación estadística aceptamos la Ha; además, al ser la probabilidad asociada al Test de ADF menor a 0.05 se corrobora la aceptación de la Ha, y se concluye que no existe raíz unitaria, por tanto, la serie es estacionaria.

<u>Conclusión de las Pruebas DF y ADF:</u> Derivado de los resultados obtenidos y en virtud de que con ambas pruebas se acepta la Ho con los datos en niveles, pero se rechaza la Ho en primeras diferencias, entonces la serie contiene una raíz unitaria y, por tanto, es no estacionaria y es integrada de orden 1, I(1). Por lo que, de acuerdo a los resultados obtenidos, es necesario diferenciar una vez la serie para volverla estacionaria

g) <u>Test PP.</u>

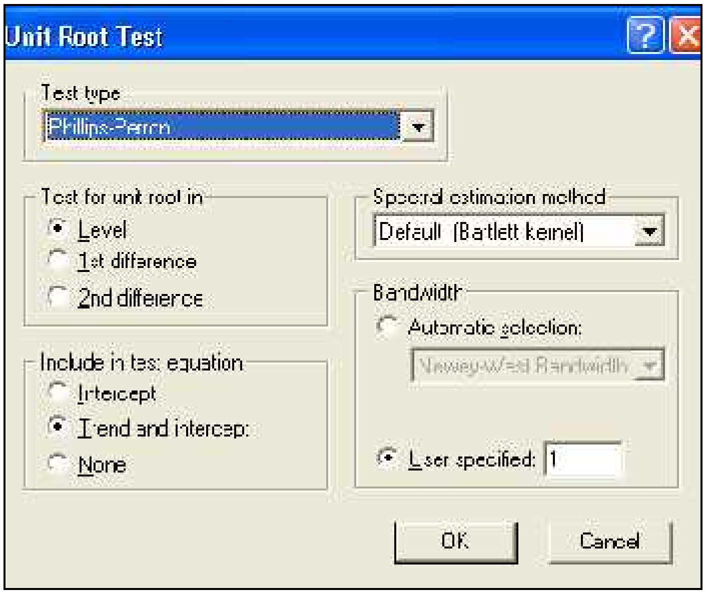
Este Test es similar a la prueba de ADF, pues si el valor del estadístico PP es menor o inferior a los valores críticos de MacKinnon, a cualquier nivel de significación en términos absolutos en ambos casos, se acepta la Ho de existencia de una raíz unitaria.

Mientras que el Test de ADF corrige la correlación serial de orden elevado añadiendo más retardos del término diferenciado de la serie original en el lado derecho de la ecuación, el Test de PP realiza una corrección del estadístico “τ” sobre el coeficiente en la regresión AR(1) para considerar la correlación serial en el término “e”.

En este Test, al igual que en el ADF, tenemos que especificar si incluimos una constante, constante más término de tendencia lineal o nada en la regresión. Para este Test hay que especificar el número de periodos de correlación serial a incluir, es decir, rezagos o retardos Pulido y López, 1999: 271).

En Eviews estando en la ventana de la serie vamos a: View/Unit Root Test… (test de raíces unitarias)/, en donde seleccionaremos el Test de Phillips-Perron (PP), en “test for unit root in” seleccionamos primeramente niveles (Level), en “Include in test equation” seleccionamos “Trend and intercept”, en “Spectral estimation method” vamos a dejar el que por default nos proporciona Eviews, es decir, el Bartlett Kernel, y en “Bandwidth” vamos a seleccionar especificado por el usuario (User specified) en donde escribiremos 1 y Ok. De tal suerte que, la ventana de raíces unitarias con el Test de PP debe visualizarse de la siguiente manera:

Cuadro 12.4.3.10.: Raíz Unitaria de la Prueba de Phillips-Perron.



El resultado del Test de PP en niveles, con intercepto y término de tendencia lineal y bandwidth especificado por el usuario de 1 es:

Cuadro 12.4.3.11.: Análisis de Regresión de la Prueba de Phillips-Perron.

Null Hypothesis: ARA has a unit root					
Exogenous: Constant, Linear Trend					
Bandwidth: 1 (Fixed using Bartlett kernel)					
			Adj. t-Stat		**Prob.***
Phillips-Perron test statistic			**-2.485894**		**0.3350**
Test critical values:	**1% level**		**-3.990701**		
	5% level		**-3.425728**		
	10% level		**-3.136027**		
*MacKinnon (1996) one-sided p-values.					
Residual variance (no correction)					0.056955
HAC corrected variance (Bartlett kernel)					0.059988
Phillips-Perron Test Equation					
Dependent Variable: D(ARA)					
Method: Least Squares					
Date: 03/02/09 Time: 16:54					
Sample(adjusted): 2 285					
Included observations: 284 after adjusting endpoints					
Variable	Coefficient	Std. Error	t-Statistic		Prob.
ARA(-1)	-0.036564	0.014995	-2.438438		0.0154

C	0.466007	0.199219	2.339171	0.0200
@TREND(1)	-0.001323	0.000536	-2.470401	0.0141

R-squared	0.021590	Mean dependent var	-0.027254
Adjusted R-squared	0.014627	S.D. dependent var	0.241696
S.E. of regression	0.239922	Akaike info criterion	-0.006496
Sum squared resid	16.17512	Schwarz criterion	0.032050
Log likelihood	3.922429	F-statistic	3.100400
Durbin-Watson stat	1.890580	Prob(F-statistic)	0.046575

<u>Comentario:</u> Como el valor del estadístico PP es, es términos absolutos, menor que el valor crítico de MacKinnon al 5% de significación estadística (2.48 < 3.42) se acepta la Ho y dado que la probabilidad asociada al estadístico PP es mayor a 0.05 se acepta la Ho de que existe raíz unitaria, por tanto, la serie es no estacionaria.

Ahora, realicemos la prueba PP en primeras diferencias manteniendo constante los demás términos de la selección, así obtenemos:

Cuadro 12.4.3.12.: Análisis de Regresión de la Prueba de Phillips-Perron.

Null Hypothesis: D(ARA) has a unit root
Exogenous: Constant, Linear Trend
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		**-16.12566**	**0.0000**
Test critical values:	**1% level**	**-3.990817**	
	5% level	**-3.425784**	
	10% level	**-3.136061**	

*MacKinnon (1996) one-sided p-values.

Residual variance (no correction)	0.058116
HAC corrected variance (Bartlett kernel)	0.058396

Phillips-Perron Test Equation
Dependent Variable: D(ARA,2)
Method: Least Squares
Date: 03/02/09 Time: 16:59
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(ARA(-1))	-0.962111	0.059669	-16.12400	0.0000
C	-0.017313	0.029055	-0.595879	0.5517
@TREND(1)	-6.78E-05	0.000176	-0.384186	0.7011

R-squared	0.481487	Mean dependent var	-0.000353
Adjusted R-squared	0.477784	S.D. dependent var	0.335380
S.E. of regression	0.242361	Akaike info criterion	0.013768
Sum squared resid	16.44689	Schwarz criterion	0.052412
Log likelihood	1.051834	F-statistic	130.0031
Durbin-Watson stat	1.984524	Prob(F-statistic)	0.000000

Comentario: Dado que el valor del estadístico PP es mayor, en términos absolutos, que el valor crítico de MacKinnon al 5% (16.12 > 3.42) y en virtud de que la probabilidad asociada al estadístico es menor a 0.05, aceptamos la Ha y rechazamos la Ho, por tanto, no existe raíz unitaria, es decir, la serie es estacionaria.

Conclusión del Test PP: Dado que con la prueba PP en niveles se acepta la Ho y en primeras diferencias se acepta la Ha, concluimos que la serie original "ARA" no es estacionaria y por tanto hay que diferenciarla una vez para convertirla en estacionaria.

Así, tanto los métodos informales como los formales nos indican que la serie original "ARA" es no estacionaria, por lo que es necesario diferenciarla una vez para volverla estacionaria. Luego entonces, la serie es integrada de orden 1, I(1). Asimismo, es preciso, antes de diferenciarla, convertirla en logaritmos en virtud de la dispersión que presentan algunas variables aleatorias, lo cual es indicio de la inexistencia de varianza constante.

VII) Transformación y diferenciación de la serie.

Teóricamente, para obtener una serie en media constante es preciso obtener sus primeras diferencias y, para obtener una serie con varianza constante es necesario transformarla en logaritmos antes de diferenciarla, ya que las series transformadas en logaritmos presentan menor dispersión.

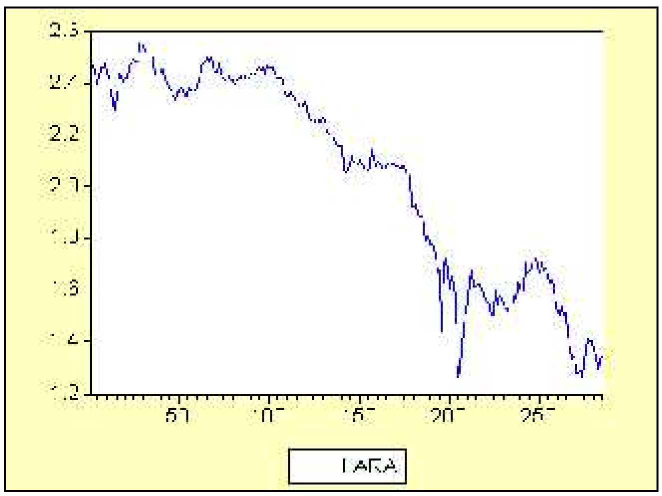
Recuérdese que el orden de integración "I" es el número de operaciones de diferenciación que hay que efectuar para convertir la serie en estacionaria (Pulido y López, 1999: 269).

De esta forma, primero debemos generar la serie en logaritmos y graficarla de forma lineal y a través del diagrama de dispersión para ver si disminuyó la dispersión que presenta. Así, vamos al menú principal y seleccionamos: Quick/Generate Series…/, y en la ventana que nos desplegará escribiremos: lara=log(ara)/Ok., donde "l" indica que la serie fue transformada en logaritmos, ello se observa de la siguiente manera:

Cuadro 12.4.3.13.: Creación de la variable ARA con logaritmos.

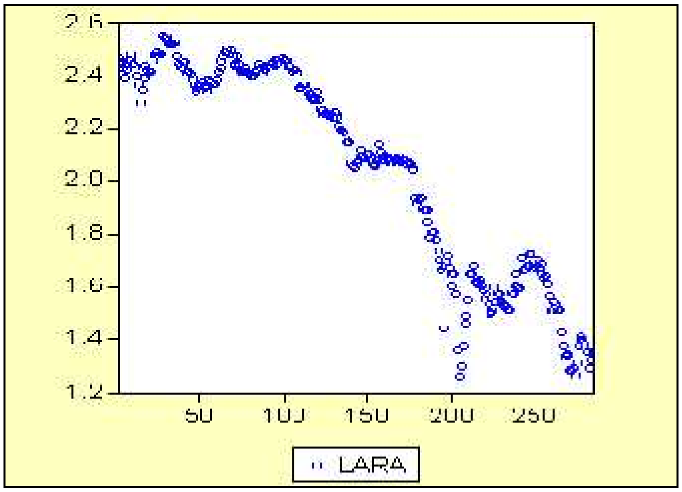
La nueva variable generada (lara) aparecerá en la ventana del Workfile. A continuación procedemos a graficarla de forma lineal a través de los pasos ya descritos anteriormente. De esta forma obtenemos:

Gráfica 12.4.3.3.: Comportamiento logarítmico de “ARA”- “LARA”



Su diagrama de dispersión es:

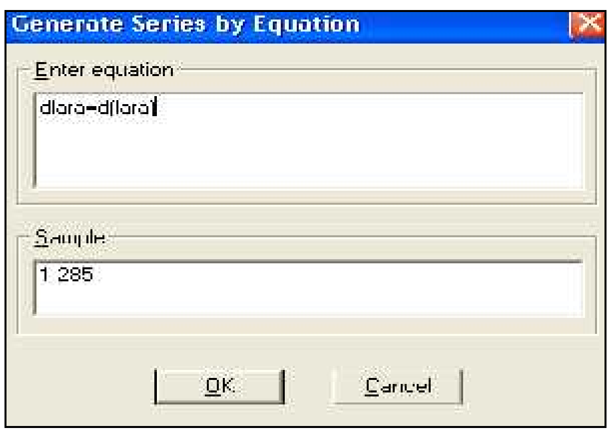
Gráfica 12.4.3.4.: Gráfico de dispresión logarítmico de “ARA”



Como puede observarse, la transformación en logaritmos de la serie (lara) aun presenta un comportamiento similar al de la serie original (ara), es decir, sin transformar. Diferenciemos entonces la serie transformada en logaritmos una vez y grafiquémosla para ver su comportamiento.

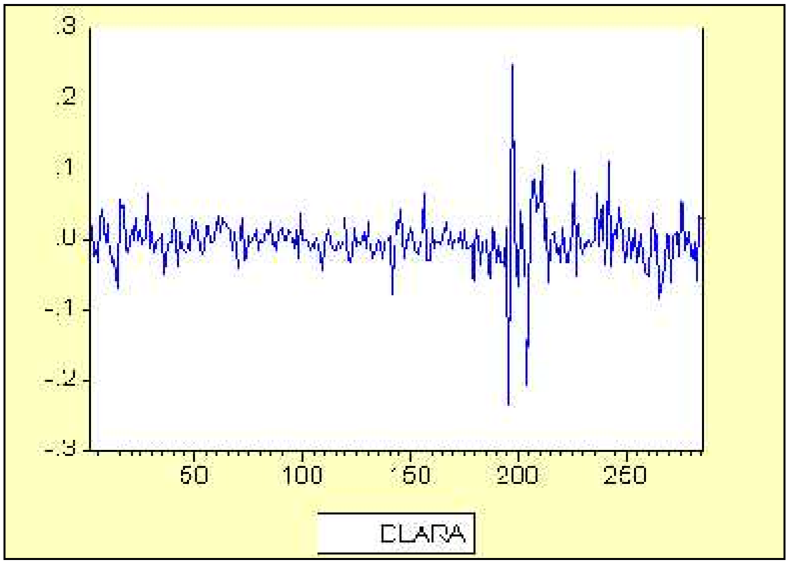
El procedimiento para diferenciar la nueva serie en logaritmos (lara) es el siguiente, en el menú principal vamos a: Quick/Generate Series…/, en la ventana que desplegará escribiremos dlara=d(lara)/Ok., donde "d" indica que se trata de la primera diferencia de la serie "lara". Ello se observa así:

Cuadro 12.4.3.14.: Creación de la diferencial de la variable "LARA"



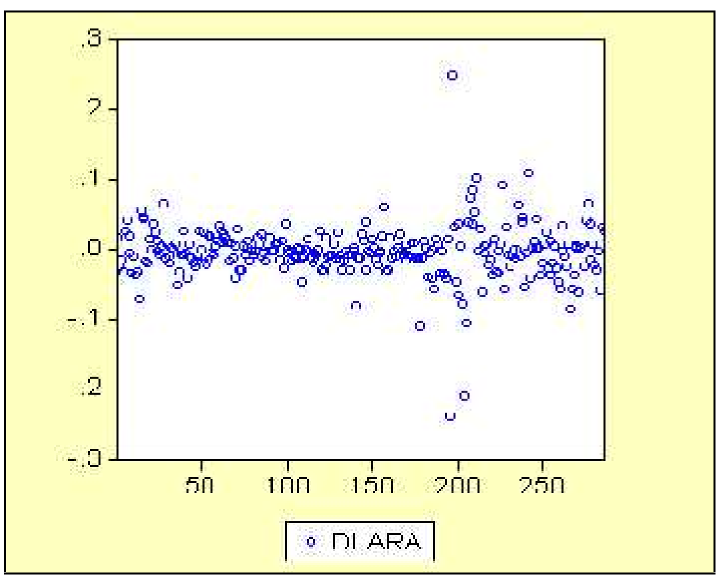
La nueva serie en logaritmos y diferenciada deberá aparecer en la ventana del Workfile. Una vez ahí, vamos a abrir la serie "dlara" dando doble click sobre la misma y a continuación vamos a graficarla a través del procedimiento conocido, es decir, en la ventana de la serie vamos a: View/Graph/Line/Ok., lo que nos desplegará la gráfica siguiente:

Gráfica 12.4.3.5.: Comportamiento de la variable "DLARA"



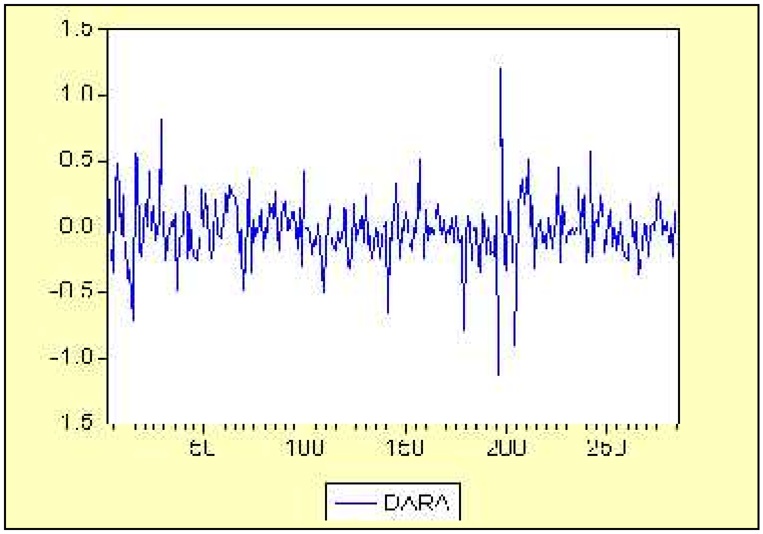
Como puede apreciarse, la serie "dlara" presenta gráficamente un comportamiento paralelo al eje de las abscisas, lo que indica una media constante y la no existencia de tendencia. Asimismo, el diagrama de dispersión nos muestra que la mayoría de las variables aleatorias oscilan alrededor de la media como puede observarse en el diagrama de dispersión de "dlara"

Gráfica 12.4.3.6.: Dispersión "DLARA"



Pero, ¿Será adecuada la transformación y diferenciación correspondiente que realizamos? La forma de saberlo es crear directamente la serie "ara" en diferencias y obtener su gráfico correspondiente. Para ello, vamos al menú principal y ahí seleccionamos: Quick/Generate Series…/, y en la ventana que nos desplegará escribimos dara=d(ara). De la cual su gráfico correspondiente es:

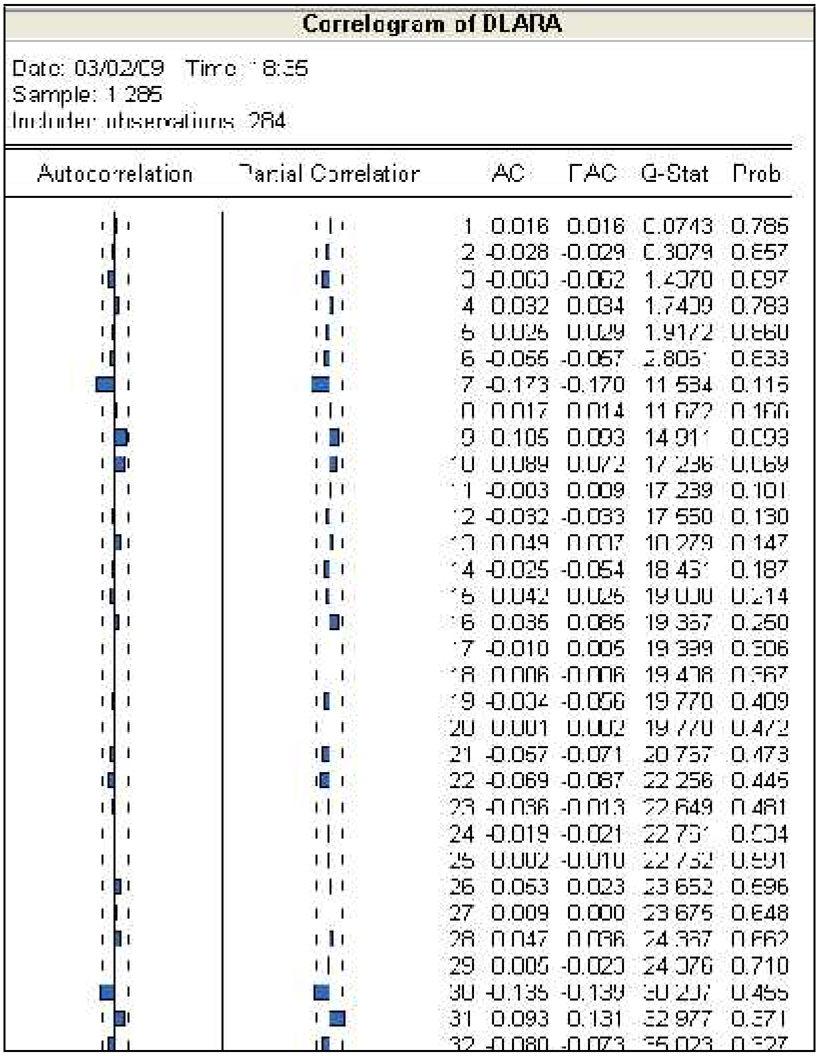
Gráfica 12.4.3.7.: Comportamiento en Diferencias de “DARA”



Como puede observarse, en la gráfica de la serie “dlara” los valores de las variables aleatorias oscilan en torno al 0.2 y al -0.2, mientras que en la serie “dara” dichos valores oscilan en torno al 1.0 y al -1.0, por tanto, los valores de la serie transformada en logaritmos y diferenciada una vez presentan menor dispersión, lo que supone una media y varianza constante. Entonces, para nuestro caso no cabe duda de la serie con la que vamos a trabajar, es decir, aquella que fue transformada en logaritmos y diferenciada una vez (dlara) pues, es estacionaria.

Con ello, podemos ahora realizar el correlograma de la serie “dlara” para verificar la estacionariedad de la misma. Así, con los pasos ya conocidos obtenemos el siguiente correlograma:

Cuadro 12.4.3.15.: Correlograma.



Correlogram of DLARA

Date: 03/02/09 Time: 18:35
Sample: 1 285
Included observations: 284

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.016	0.016	0.0743	0.785
		2	-0.028	-0.029	0.3079	0.857
		3	-0.063	-0.062	1.4370	0.697
		4	0.032	0.034	1.7409	0.783
		5	0.026	0.029	1.9172	0.860
		6	-0.056	-0.057	2.8051	0.833
		7	-0.173	-0.170	11.534	0.116
		8	0.017	0.014	11.672	0.166
		9	0.105	0.093	14.911	0.093
		10	0.089	0.072	17.236	0.069
		11	-0.003	0.009	17.239	0.101
		12	-0.032	-0.033	17.550	0.130
		13	0.049	0.007	18.279	0.147
		14	-0.025	-0.054	18.431	0.187
		15	0.042	0.025	19.000	0.214
		16	0.035	0.085	19.357	0.250
		17	-0.010	0.005	19.389	0.306
		18	0.006	-0.006	19.478	0.367
		19	-0.034	-0.056	19.770	0.409
		20	0.001	0.002	19.770	0.472
		21	-0.057	-0.071	20.757	0.473
		22	-0.069	-0.087	22.256	0.445
		23	-0.036	-0.013	22.649	0.481
		24	-0.019	-0.021	22.751	0.534
		25	0.002	-0.010	22.752	0.591
		26	0.053	0.023	23.652	0.596
		27	0.009	0.000	23.675	0.648
		28	0.047	0.036	24.337	0.662
		29	0.005	-0.020	24.376	0.710
		30	-0.135	-0.139	30.207	0.455
		31	0.093	0.131	32.977	0.371
		32	-0.080	-0.073	35.023	0.327

Como puede observarse, al encontrarse los valores de la autocorrelación (AC) y los de la autocorrelación parcial (PAC) dentro de las bandas de confianza, al ser los valores de AC y PAC muy menores a 1 y al ser la probabilidad asociada al estadístico “Q” mayores a 0.05 concluimos, de acuerdo a este correlograma, que la serie “dlara” es estacionaria.

<u>Ahora, realicemos los Test de raíces unitarias correspondientes, es decir, las pruebas o métodos formales, para determinar que la serie efectivamente es ya estacionaria, pues, gráficamente lo es.</u>

g) <u>Test de DF para “dlara”.</u>

Siguiendo los pasos ya descritos con antelación para realizar la prueba DF y, planteando la prueba de hipótesis correspondiente de la siguiente manera:

Ho: ρ = 1; existe raíz unitaria, por tanto, la serie es no estacionaria.

Ha: ρ < 1; no existe raíz unitaria, por tanto, la serie es estacionaria.

Cabe hacer mención que para realizar la prueba DF va a hacer en niveles (Level) porque ya está diferenciada una vez, por lo que si se selecciona primeras diferencias en realidad se tratará de las segundas diferencias; dado que la nueva serie (dlara) no parece presentar tendencia (creciente o decreciente) pero parece tener uno o más valores medios distintos de cero debemos incluir un término constante (intercept) en la regresión; y en el número de rezagos cero para realizar el Test DF. El resultado del Test de DF es:

Cuadro 12.4.3.16.: Análisis de Regresión de la Prueba de Dickey-Fuller Aumentada.

Null Hypothesis: DLARA has a unit root
Exogenous: Constant
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-16.48547**	**0.0000**
Test critical values:	**1% level**	**-3.453317**	
	5% level	**-2.871546**	
	10% level	**-2.572174**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DLARA)
Method: Least Squares
Date: 03/02/09 Time: 18:58
Sample(adjusted): 3 285
Included observations: 283 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLARA(-1)	-0.983876	0.059681	-16.48547	0.0000
C	-0.003884	0.002284	-1.700752	0.0901

R-squared	0.491652	Mean dependent var	3.23E-05
Adjusted R-squared	0.489843	S.D. dependent var	0.053500
S.E. of regression	0.038213	Akaike info criterion	-3.684263
Sum squared resid	0.410316	Schwarz criterion	-3.658500
Log likelihood	523.3232	F-statistic	271.7706
Durbin-Watson stat	1.993651	Prob(F-statistic)	0.000000

Como el valor del estadístico DF (16.48) en términos absolutos es mayor que el valor crítico de MacKinnon al 5% (2.87) y dado que la probabilidad del estadístico DF es menor a 0.05 se acepta la Ha y se rechaza la Ho, por tanto, no existe raíz unitaria, es decir, la serie “dlara” es estacionaria.

h) <u>Test de ADF para “dlara”.</u>

Ahora realicemos la prueba ADF, manteniendo los términos seleccionados constantes pero modificando el número de rezagos a 1 en lugar de cero. Obtenemos:

Cuadro 12.4.3.17.: Análisis de Regresión de la Prueba de Dickey-Fuller Aumentada.

Null Hypothesis: DLARA has a unit root
Exogenous: Constant
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		**-12.01737**	**0.0000**
Test critical values:	**1% level**	**-3.453400**	
	5% level	**-2.871582**	
	10% level	**-2.572193**	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DLARA)
Method: Least Squares
Date: 03/02/09 Time: 19:03
Sample(adjusted): 4 285
Included observations: 282 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLARA(-1)	-1.011255	0.084149	-12.01737	0.0000
D(DLARA(-1))	0.028867	0.059940	0.481598	0.6305
C	-0.003922	0.002307	-1.699698	0.0903

R-squared	0.491521	Mean dependent var	0.000181
Adjusted R-squared	0.487876	S.D. dependent var	0.053536
S.E. of regression	0.038312	Akaike info criterion	-3.675520
Sum squared resid	0.409521	Schwarz criterion	-3.636777
Log likelihood	521.2484	F-statistic	134.8477
Durbin-Watson stat	2.002513	Prob(F-statistic)	0.000000

En virtud de que el valor del estadístico ADF (12.01) es mayor que el valor crítico de MacKinnon al 5% (2.87) y dado que la probabilidad asociada al estadístico ADF es menor a 0.05 se acepta la Ha, por tanto, no existe raíz unitaria, es decir, la serie “dlara” es estacionaria.

i) <u>Test de PP para “dlara”.</u>

Para el caso del Test de PP vamos a realizarlo en niveles (Level); con intercepto (intercept); y vamos a especificar de forma manual 1 rezago. Así obtenemos lo siguiente:

Cuadro 12.4.3.18.: Análisis de Regresión de la Prueba de Dickey-Fuller Aumentada.

Null Hypothesis: DLARA has a unit root				
Exogenous: Constant				
Bandwidth: 1 (Fixed using Bartlett kernel)				
			Adj. t-Stat	**Prob.***
Phillips-Perron test statistic			**-16.48574**	**0.0000**
Test critical values:	**1% level**		**-3.453317**	
	5% level		**-2.871546**	
	10% level		**-2.572174**	
*MacKinnon (1996) one-sided p-values.				
Residual variance (no correction)				0.001450
HAC corrected variance (Bartlett kernel)				0.001452

Phillips-Perron Test Equation				
Dependent Variable: D(DLARA)				
Method: Least Squares				
Date: 03/02/09 Time: 19:11				
Sample(adjusted): 3 285				
Included observations: 283 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DLARA(-1)	-0.983876	0.059681	-16.48547	0.0000
C	-0.003884	0.002284	-1.700752	0.0901
R-squared	0.491652	Mean dependent var		3.23E-05
Adjusted R-squared	0.489843	S.D. dependent var		0.053500
S.E. of regression	0.038213	Akaike info criterion		-3.684263
Sum squared resid	0.410316	Schwarz criterion		-3.658500
Log likelihood	523.3232	F-statistic		271.7706
Durbin-Watson stat	1.993651	Prob(F-statistic)		0.000000

Como puede observarse en los resultados del Test de PP el estadístico PP es mayor (16.48) al valor crítico de MacKinnon al 5% (2.87) y dado que la probabilidad asociada al estadístico PP es menor a 0.05, entonces aceptamos la Ha y rechazamos la Ho, por lo que no existe raíz unitaria, luego entonces la serie “dlara” es estacionaria.

***Conclusión general*:** Dado que con los métodos informales (gráfica lineal, diagrama de dispersión y correlograma) concluimos que la serie “dlara”, misma que fue transformada en logaritmos y diferenciada una vez, es estacionaria; y con los métodos formales (Test de DF, ADF y PP) se acepta la Ha en virtud de que los valores de los estadísticos DF, ADF y PP son mayores a los valores críticos de MacKinnon al 5% y a que la probabilidad asociada a los mismos es menor a 0.05 *concluimos contundentemente que la serie de tiempo “dlara” es estacionaria.* Por tanto, podemos pasar a la identificación del modelo ARMA, pues ya sabemos que la serie es integrada de orden 1, es decir, I(1).

VIII) Identificación de la estructura ARMA.

Una vez que hemos transformado la serie en estacionaria o en su caso probado que la serie es estacionaria de origen procedemos a determinar el tipo de modelo más adecuado para la serie objeto de estudio, es decir, el orden de los procesos autorregresivos (AR) y de medias móviles (MA), esto es, los valores de p y q.

Sin embargo, no se piense que resulta fácil, incluso para personas con experiencia, seleccionar el modelo ARMA. En muchas ocasiones se tendrán dudas razonables sobre el modelo más adecuado. Al respecto, Pulido y López (1999: 294, 301) sostienen que "no debemos preocuparnos en exceso por ello, ya que una posibilidad es seleccionar varios modelos alternativos y comprobar posteriormente cuál de ellos resulta finalmente más idóneo. Generalmente, suele procederse seleccionando siempre procesos lo más sencillos posible". Asimismo, Johnston y Dinardo (2001) reconocen lo difícil que resulta la elección del orden de p y q. No obstante, sostienen que esta elección se agiliza mediante un procedimiento numérico sugerido por Hannan y Rissanen. Se trata de un procedimiento en tres etapas, a saber:

1.- Se estiman por MCO procesos AR puros de orden bastante elevado, lo cual no deja de ser razonable ya que un proceso ARMA desconocido equivale a un proceso AR infinito;

2.- Se selecciona la regresión que ha proporcionado el menor valor según el criterio de información de Akaike (AIC) y los residuos de dicha regresión se consideran como estimadores de los errores de un modelo ARMA; y

3.- Se ajustan algunos modelos ARMA utilizando para ello los residuos estimados. Después del ajuste, es decir, de la inclusión de MA(q) elegimos como especificación correcta aquella que proporciona el menor valor del criterio de Schwarz.

Técnicamente, la decisión de que proceso a incluir en el modelo y su orden se tomará a partir de las funciones de autocorrelación (AC) y autocorrelación parcial (PAC). Es decir, el instrumento técnico básico para identificar un modelo ARMA es la denominada función de autocorrelación, misma que mide el grado de correlación entre cada valor de la variable y los desfasados 1, 2, …, h periodos (Pulido y López, 1999: 285-286). Habitualmente se terminará eligiendo entre los procesos más simples: AR(1), AR(2), MA(1), MA(2) y ARMA(1,1). En caso de duda pueden seleccionarse varios modelos alternativos, que serán estimados y contrastados posteriormente, para decidir el definitivamente adoptado.

Por su parte, para identificar un proceso AR, si la función de autocorrelación decrece (en forma regular, es decir, alternando valores positivos y negativos o en forma de ondas sinusoidales) y la función de autocorrelación parcial presenta un número de coeficientes igual al orden del proceso, entonces, el AR corresponderá a el número de coeficientes de la autocorrelación parcial que sobrepasan las bandas de confianza.

Un modelo MA(1) tiene sólo el primer coeficiente distinto de cero en la función de autocorrelación y la función de autocorrelación parcial muestra un comportamiento decreciente; un modelo MA(2) tendrá dos coeficientes no nulos en la función de autocorrelación mientras que la función de autocorrelación parcial mostrará un comportamiento decreciente.

Un modelo ARMA(1,1) presentará diversas combinaciones, pero siempre mostrará un comportamiento decreciente en ambas funciones, es decir, la de autocorrelación y la de autocorrelación parcial.

Por tanto, a la hora de identificar un modelo ARMA debemos siempre observar las funciones de autocorrelación para decidir el orden del componente regular (p,q) con base en la evolución de los primeros coeficientes.

En el cuadro 1 se incluyen los procesos AR, MA y ARMA que deberán utilizarse en caso de que la serie estacionaria muestre un comportamiento en la función de autocorrelación y autocorrelación parcial similar a las funciones teóricas de autocorrelación y autocorrelación parcial contenidas en dicho cuadro.

Otro criterio proporcionado por Pulido y López (1999: 291) para determinar los procesos a incluir en el modelo es el siguiente:

Cuadro 12.4.3.19.: Funciones de autocorrelación.

Función de autocorrelación parcial / **Función de autocorrelación**	Decrece	Uno o dos coeficientes salen de las bandas de confianza (son significativos)
Decrece	ARMA(1,1)	AR(1) o AR(2)
Uno o dos coeficientes salen de las bandas de confianza (son significativos)	MA(1) o MA(2)	No posible

Finalmente, recuérdese y téngase presente que la forma general de un AR(p) es:

$$Y_t = C + ¤_1 Y_{t-1} + ¤_2 Y_{t-2} + ... + ¤_p Y_{t-p} + a_t$$

Por tanto, un AR(1) tiene la forma:

$$Y_t = C + ¤_1 Y_{t-1} + a_t$$

Mientras que un AR(2):

$$Y_t = C + ¤_1 Y_{t-1} + ¤_2 Y_{t-2} + a_t$$

Por su parte, la forma general de MA(q) es:

$$Y_t = « + a_t - ¯_1 a_{t-1} - ¯_2 a_{t-2} - ... - ¯_q a_{t-q}$$

Donde a_{t-1} son los residuos.

Así, un proceso MA(1) esta dado por:

$$Y_t = « + a_t - ¯_1 a_{t-1}$$

Mientras que un MA(2) viene dado por:

$$Y_t = « + a_t - ¯_1 a_{t-1} - ¯_2 a_{t-2}$$

Y un proceso combinado general ARMA(p,q) está dado por:

$$Y_t = C + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + ... + \phi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - ... - \theta_q a_{t-q}$$

Cabe destacar, como comentan Pulido y López (1999: 281), que "realmente, en las aplicaciones no es frecuente manejar un modelo ARIMA(p,d,q), sino el producto de dos modelos: ARIMA(p,d,q)*SARIMA(P,D,Q), donde el primero corresponde a la parte regular y el segundo a la estacional de la serie.

La existencia de una tendencia en la estacionalidad (a parte de la tendencia regular de la serie) puede venir provocada por una estacionalidad no rígida en que, por ejemplo, las puntas de estacionalidad se agudicen (o reduzcan) con el paso del tiempo.

Ahora bien, para identificar los procesos que se incluirán en el modelo o los modelos posibles a estimar, como se comento, es necesario contar con las funciones de autocorrelación y autocorrelación parcial del correlograma de la serie estacionaria (ver la página 25), en nuestro caso de la serie: DLARA.

Como puede observarse en el correlograma, el rezago séptimo sale de las bandas de confianza tanto de las funciones de autocorrelación como de autocorrelación parcial; asimismo, el rezago 30 sale de las bandas de confianza para cada función; y, el rezago 31 sale de las bandas de confianza en la función de autocorrelación parcial.

Cuadro 12.4.3.19.: Comportamiento teórico para identificar en la función de autocorrelación total y parcial los procesos AR, MA Y ARMA.

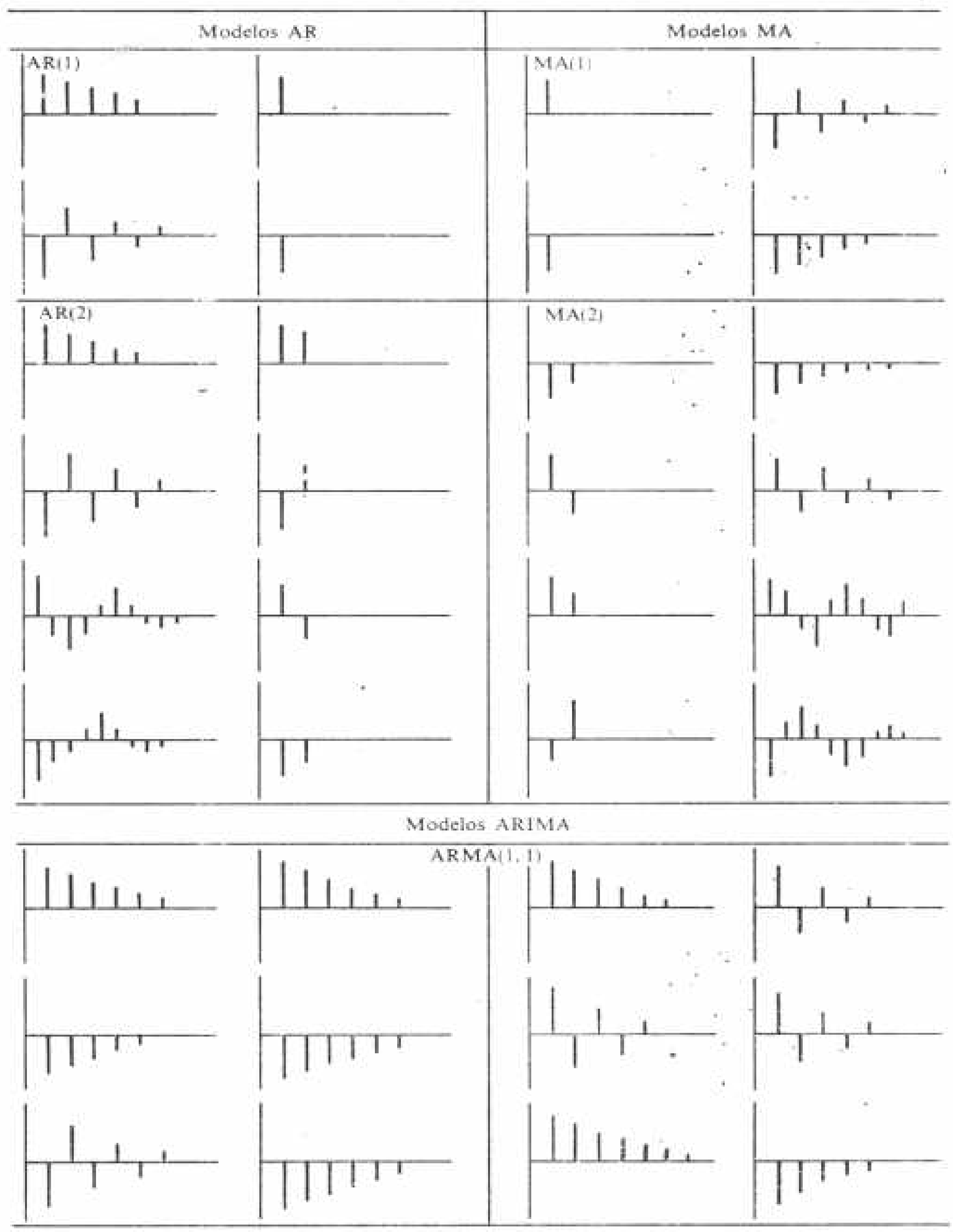


Figura 20.2.—Funciones teóricas de autocorrelación y autocorrelación parcial (izquierda y derecha en cada caso) para diferentes tipos de modelos.

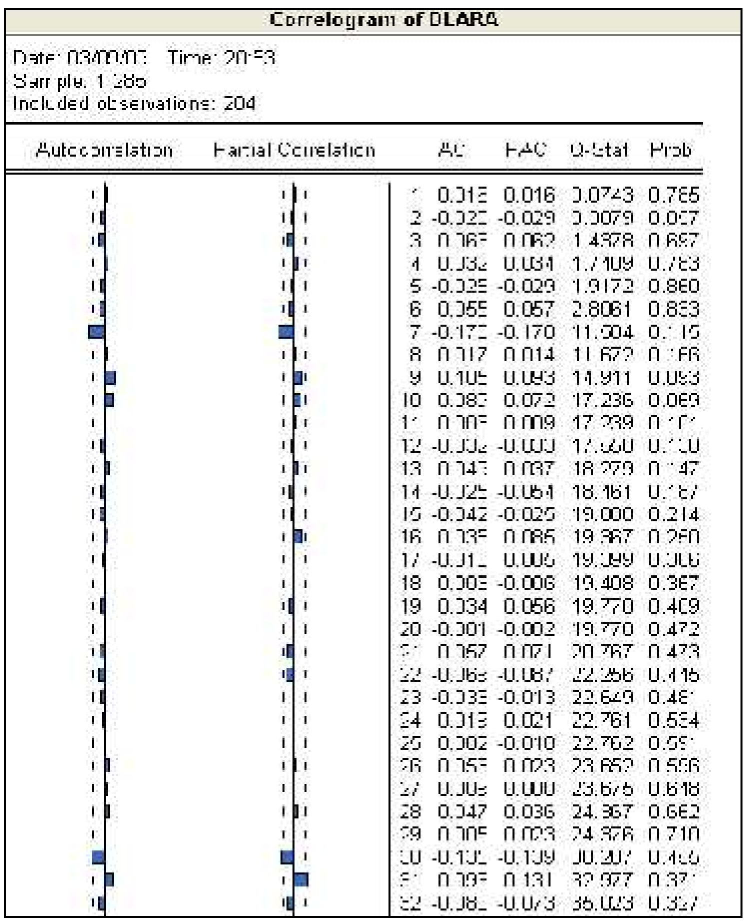
En virtud de que el comportamiento mostrado en el correlograma de la serie DLARA no corresponde con el comportamiento teórico establecido en el cuadro 1, vamos a establecer dos modelos hipotéticos para, posteriormente, de acuerdo a determinados criterios que veremos más adelante, discernir entre uno u otros modelos.

De tal suerte y derivado del análisis visual realizado a través del correlograma para el caso de la serie DLARA y tras realizar diversos modelos con ayuda de Eviews, los modelos identificados que vamos a desarrollar aquí son:

1. AR(7) MA(7), es decir, ARMA(7,7) y en virtud de que diferenciamos una vez la serie para convertirla en estacionaria es integrada de orden 1, I(1). Por tanto, nuestro primer modelo tiene la forma: ARIMA(7,1,7); y

2. AR(7) MA(7) y en la parte estacional un SMA(31), y como diferenciamos una vez para convertir la serie en estacionaria, la serie DLARA es integrada de orden 1, I(1). Por tanto, el segundo modelo a estimar y contrastar es: ARIMA(7,1,7)*SARIMA(0,0,31).

Cuadro 12.4.3.20.: Correlograma.



Correlogram of DLARA

Date: 03/09/03 Time: 20:53
Sample: 1 285
Included observations: 204

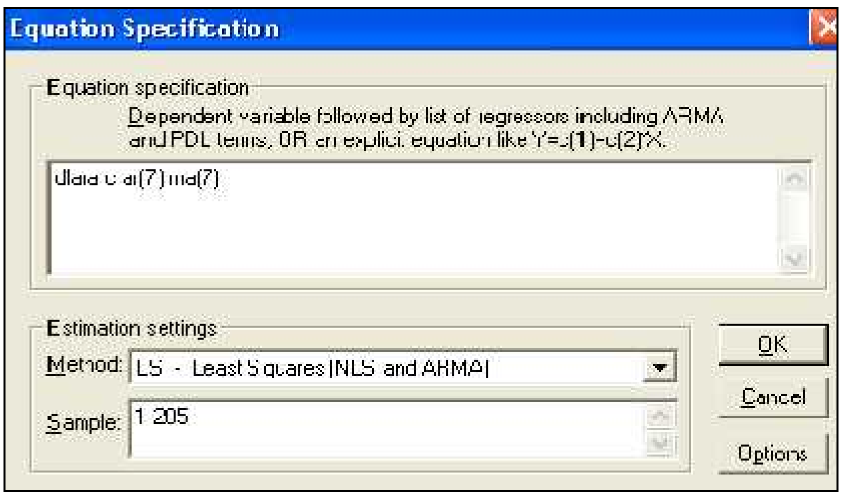
	AC	PAC	Q-Stat	Prob
1	0.013	0.016	0.0743	0.785
2	-0.023	-0.029	0.3079	0.927
3	0.067	0.062	1.4378	0.697
4	0.032	0.034	1.7109	0.789
5	-0.025	-0.029	1.9172	0.860
6	0.065	0.057	2.8061	0.833
7	-0.173	-0.170	11.504	0.118
8	0.017	0.014	11.672	0.166
9	0.105	0.093	14.911	0.093
10	0.083	0.072	17.236	0.069
11	0.005	0.009	17.239	0.101
12	-0.032	-0.003	17.460	0.130
13	0.043	0.037	18.279	0.147
14	-0.025	-0.054	18.461	0.187
15	-0.042	-0.025	19.000	0.214
16	0.035	0.085	19.367	0.250
17	-0.012	0.006	19.399	0.306
18	0.003	-0.006	19.408	0.367
19	0.034	0.056	19.770	0.409
20	-0.001	-0.002	19.770	0.472
21	0.067	0.071	20.767	0.473
22	-0.063	-0.087	22.256	0.415
23	-0.033	-0.013	22.649	0.481
24	0.013	0.021	22.761	0.534
25	0.002	-0.010	22.762	0.591
26	0.053	0.023	23.652	0.596
27	0.003	0.000	23.676	0.648
28	0.047	0.036	24.367	0.662
29	0.005	0.003	24.376	0.710
30	-0.132	-0.109	30.207	0.456
31	0.097	0.131	32.977	0.371
32	-0.082	-0.073	35.023	0.327

IX) Estimación de coeficientes y evaluación-validación del modelo.

La evaluación de modelos ARIMA puede hacerse en forma similar a como se hace con el modelo de regresión o los modelos de ecuaciones simultaneas (Pulido y Pérez, 2000)

Así, una vez identificado el modelo o los modelos procedemos a estimarlos con ayuda de Eviews de la siguiente forma: en el menú principal vamos a: Quick/Estimate Equation/, y en la ventana que despliega escribimos el primer modelo: dlara c ar(7) ma(7)/Ok.

Cuadro 12.4.3.21.: Especificación de la ecuación.



Con lo cual obtenemos los siguientes resultados:

Cuadro 12.4.3.22.: Regresión "DLARA"

Dependent Variable: DLARA
Method: Least Squares
Date: 03/11/09 Time: 13:48
Sample(adjusted): 9 285
Included observations: 277 after adjusting endpoints
Convergence achieved after 24 iterations
Backcast: 2 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.004593	0.001342	-3.421214	0.0007
AR(7)	0.653180	0.102730	6.358233	0.0000
MA(7)	-0.820713	0.078910	-10.40057	0.0000

R-squared	0.056671	Mean dependent var	-0.003994
Adjusted R-squared	0.049785	S.D. dependent var	0.038360
S.E. of regression	0.037393	Akaike info criterion	-3.723910
Sum squared resid	0.383111	Schwarz criterion	-3.684661
Log likelihood	518.7615	F-statistic	8.230282

Durbin-Watson stat	2.016104	Prob(F-statistic)		0.000338
Inverted AR Roots	.94	.59+.74i	.59 -.74i	-.21 -.92i
	-.21+.92i	-.85 -.41i	-.85+.41i	
Inverted MA Roots	.97	.61+.76i	.61 -.76i	-.22 -.95i
	-.22+.95i	-.88 -.42i	-.88+.42i	

Con los resultados obtenidos hay que juzgar la validez del modelo estimado, para ello debemos verificar:

- Contraste "t" de significación de coeficientes: Para nuestro caso, al ser la probabilidad asociada a las estadísticas "t" para cada coeficiente menores que 0.05 concluimos que los coeficientes son estadísticamente significativos.

- Matriz de covarianzas: Para acceder a ella dentro del menú de la ecuación vamos a: View/Covariance Matriz/Ok. Obteniendo:

Cuadro 12.4.3.23.: Matriz de Covarianzas

	C	AR(7)	MA(7)
C	1.80212708265e-06	-9.36143365238e-06	5.98178985982e-06
AR(7)	-9.36143365238e-06	0.0105534103408	-0.00724748850113
MA(7)	5.98178985982e-06	-0.00724748850113	0.00622684554418

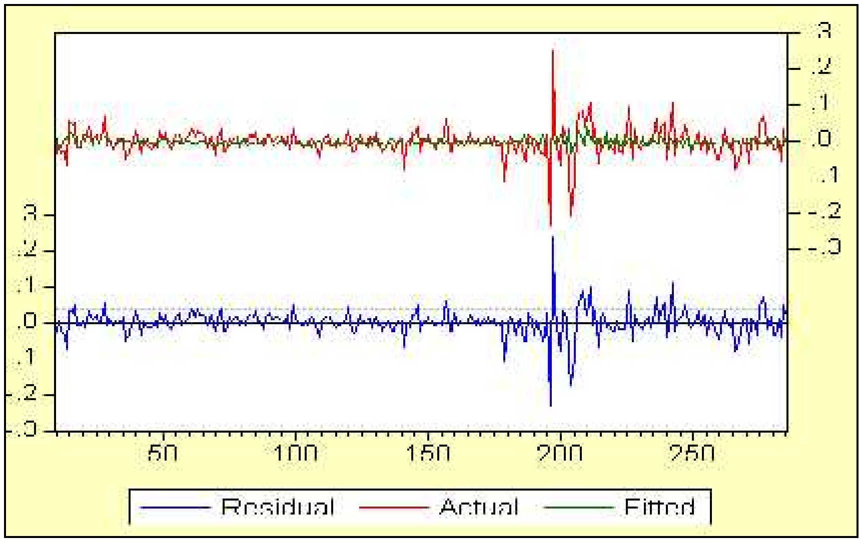
En este caso, como se recordara, lo que nos interesa analizar son los valores fuera de la diagonal principal, ¿Por qué? Así, por un lado, covarianzas relativamente altas indican parámetros en cierta forma redundantes, algunos de los cuales podrían eliminarse sin gran perjuicio para la capacidad explicativa del modelo. Por otro, covarianzas reducidas conforman la conveniencia de mantener todos los coeficientes del modelo (Pulido y López, 1999: 306).

Para nuestro caso, observamos que la covarianza que existe entre los procesos AR(7) y Ma(7) y entre cada uno de estos con la constante es muy reducida, por tanto, concluimos que debemos mantener todos los coeficientes del modelo y podemos seguir adelante.

- Tabla y gráfico de residuos: Para acceder a la tabla que nos muestra los residuos, así como los valores de la serie original y estimada, dentro del menú de la ventana de la ecuación del modelo estimado seleccionamos: View/Actual,fitted, residual table/Ok. Y para acceder el gráfico del ajuste realizado a través del modelo vamos a: View/Actual, fitted, residual graph/Ok.

Para nuestro caso la gráfica de los residuos obtenidos a través del modelo planteado es:

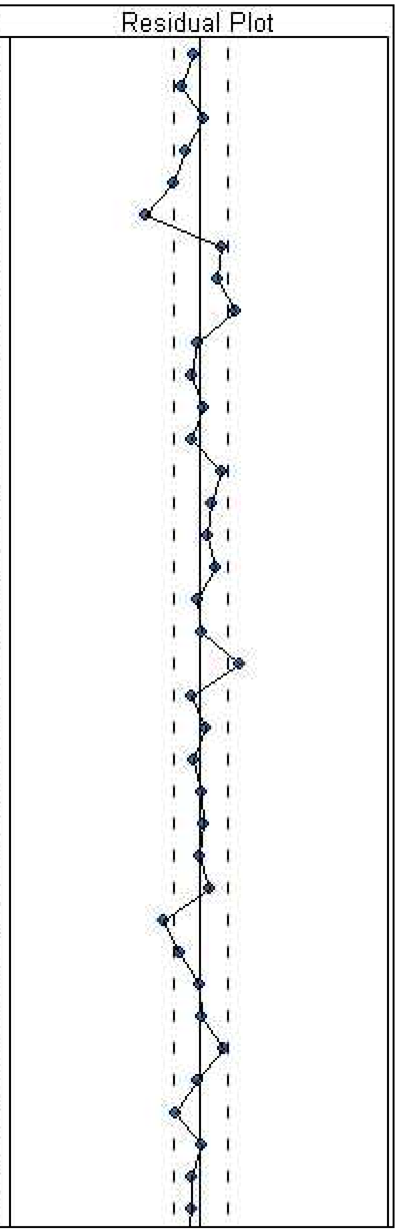
Gráfica 12.4.3.8.: Residual, Actual y Ajuste de la variable DLARA



Para el caso de la tabla de los valores reales, los estimados y la gráfica, por razones de espacio tan sólo vamos a mostrar una parte:

Gráfica 12.4.3.9.: Gráfico de parcela con ajuste residual.

obs	Actual	Fitted	Residual
9	0.02036	0.02872	-0.00836
10	-0.03416	-0.00985	-0.02432
11	-0.01048	-0.01519	0.00471
12	-0.03575	-0.01629	-0.01946
13	-0.03049	0.00483	-0.03531
14	-0.07094	0.00292	-0.07385
15	0.05486	0.02318	0.03168
16	0.04383	0.01857	0.02526
17	0.04635	-0.00395	0.05030
18	-0.01669	-0.01230	-0.00439
19	-0.01968	-0.00898	-0.01070
20	0.01435	0.00748	0.00687
21	0.00000	0.01269	-0.01269
22	0.03672	0.00824	0.02848
23	0.02375	0.00630	0.01745
24	-0.00252	-0.01260	0.01008
25	0.01253	-0.00889	0.02142
26	-0.00833	-0.00566	-0.00267
27	0.00418	0.00214	0.00204
28	0.06532	0.00882	0.05650
29	-0.01257	-0.00098	-0.01159
30	0.00787	-0.00040	0.00827
31	-0.02060	-0.01151	-0.00909
32	-0.00885	-0.01099	0.00214
33	0.00242	-0.00484	0.00726
34	-0.00081	-0.00054	-0.00027
35	0.00723	-0.00530	0.01253
36	-0.05091	-0.00029	-0.05062
37	-0.03254	-0.00324	-0.02930
38	-0.00786	-0.00759	-0.00027
39	-0.00616	-0.00913	0.00297
40	0.02699	-0.00597	0.03297
41	-0.00430	-0.00190	-0.00241
42	-0.04050	-0.00715	-0.03335
43	0.00805	0.00670	0.00136
44	-0.01165	0.00120	-0.01285
45	-0.01912	-0.00650	-0.01261



Como puede observarse en la gráfica de los valores reales, estimados y residuales, estos últimos se muestran dentro, durante la mayor parte del periodo, de las bandas de confianza, con excepción clara del periodo comprendido entre las observaciones 195 a la 215, periodo de gran volatilidad del precio de las acciones de Grupo ARA. Este periodo comprende a partir del 2 de octubre de 2008 al 6 de noviembre del mismo año, periodo que se enmarca dentro del derrumbe de las bolsas de valores del mundo, es decir, el comportamiento se explica por la incertidumbre y pánico generado a raíz del crack bursátil y por tanto quiebra de diferentes empresas a nivel mundial, entre ellas, por supuesto, las grandes compañías hipotecarias, que sin duda, impactaron desfavorablemente a diferentes empresas del ramo a nivel mundial, entre ellas: Consorcio ARA. De hecho, fue el 23 de octubre de 2008 (observación 204 correspondiente) cuando el precio de las acciones de ARA alcanzaron su precio más bajo durante todo el periodo en cuestión, a saber: 3.52 pesos por acción. Aunque, también es cierto que en las observaciones 196 y 197 el precio real de las acciones de ARA difiere en sobremanera con el precio estimado, de ahí el comportamiento que muestra o presenta la gráfica. Es decir, el error en estas observaciones es alto.

> Cabe destacar que el gráfico de los residuos nos proporciona una visión de conjunto de la cuantía de los errores, sesgos sistemáticos y puntos de errores excepcionales (Pulido y López, 1999: 311).

- Prueba F global: Como se sabe la prueba F es un contraste de significación de conjunto de todo el modelo. Así, si la probabilidad asociada al estadístico F es menor que 0.05 se concluye que las variables independientes en su conjunto explican el comportamiento de la variable dependiente y viceversa. En nuestro caso, al ser la probabilidad asociada a F menor que 0.05 concluimos que el modelo en su conjunto es adecuado.

- Coeficiente de determinación (R^2): En el análisis de series de tiempo es suficiente que el R^2 supere ampliamente los valores de 0.2 o 0.4 siempre y cuando se trabaje con un proceso estocástico en diferencias. De acuerdo a Pulido y López (1999: 311) "un coeficiente de determinación de 0.5 a 0.7 en una variable en diferencias puede ser frecuentemente equivalente o superior a 0.9 en la variable original".

En nuestro caso, tenemos que $R^2 = 0.056671$, lo cual, no cumple con lo especificado por Pulido y López; sin embrago, téngase presente que ellos hablan tan sólo de variables en diferencias, mientras que nosotros estamos trabajando la serie en logaritmos diferenciada.

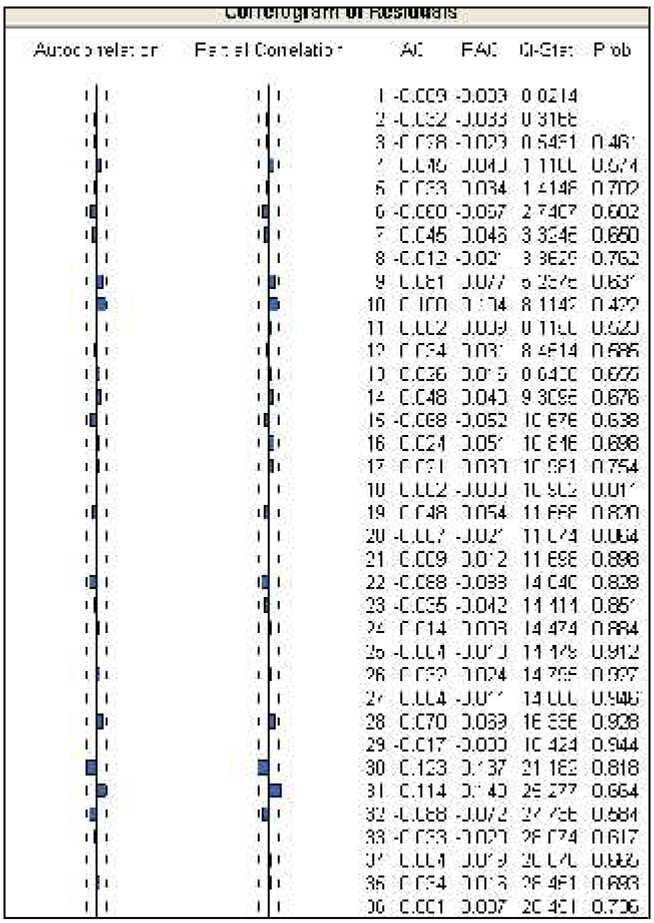
- Correlograma de los residuos: Siguiendo con la validación del modelo, vamos a observar el correlograma de los residuos obtenidos a partir del modelo planteado. Así, estando en la ventana de la ecuación vamos a: View/Residual Tests/Correlogram-Q-statistics/, mismo que por default nos da 36 rezagos/Ok., obteniendo el siguiente correlograma (ver página 30).

Como puede observarse en el correlograma de los residuos, en el rezago 7, tanto en la función de autocorrelación (AC) como en la función de autocorrelación parcial (PAC), ya no salen de las bandas de confianza. Asimismo, en la AC el rezago 30 ya no sale de las bandas de confianza, de hecho se encuentra en el límite. Sin embargo, en la PAC tanto el rezago 30 como 31 salen, aunque poco, de las bandas de confianza. Por otra parte, el valor de los coeficientes de AC y PAC son muy menores a 1. Pero como se comento, los valores para los rezagos 30 y 31 son los más altos de los 36 rezagos mostrados en el correlograma. A pesar de ello, vemos que la probabilidad asociada al estadístico Q son todas mayores a 0.05.

Derivado del análisis realizado al correlograma podemos concluir por adelantado que los residuos son ruido blanco, es decir, $\hat{\rho}_k$ ~ N(0, 1/n).

- <u>Test DF, ADF y PP para los residuos:</u> Para verificar que los residuos son ruido blanco, podemos ahora realizar sobre ellos los Test DF, ADF Y PP. Para ello, debemos crear la serie de los residuos. Así, volvemos a correr la regresión, para ello tenemos dos opciones: 1) Si estamos en la ventana de la ecuación vamos a: Estimate/Ok.; y 2) En el menú principal vamos a: Quick/Estimate Equation…/y escribimos la ecuación planteada con antelación. El porque hay que realizar estos pasos se debe a que el programa tan sólo guarda la información en tanto que no se realice una nueva operación, como es el caso de una nueva regresión, es decir, guarda la información momentáneamente. De tal suerte que, al correr la regresión o al volverla a estimar Eviews generará los residuos de la misma automáticamente en la serie “resid” que aparece en el Workfile. A continuación en el menú del Workfile vamos a: Objects/New Object…/Series/, y le damos el nombre de U1/Ok. Misma que aparecerá en la ventana del archivo de trabajo. De la serie “Resid” generada momentáneamente copiamos los valores de los residuos a partir de la observación 2 hasta la 285 y las copiamos en la nueva serie creada denominada U1 habiendo oprimido con antelación el botón “Edit +/-“ de la ventana de la nueva serie para que nos permita incluir los datos. Una vez hecho esto, podemos ahora realizar las pruebas de raíces unitarias sobre los residuos para probar si son o no ruido blanco.

Cuadro 12.4.3.24.: Correlograma de Residuales.



Correlogram of Residuals

	AC	PAC	Q-Stat	Prob
1	-0.009	-0.009	0.0214	
2	-0.032	-0.033	0.3166	
3	-0.038	-0.023	0.5451	0.461
4	0.046	0.040	1.1100	0.574
5	0.033	0.034	1.4146	0.702
6	-0.060	-0.057	2.7407	0.602
7	0.045	0.043	3.3346	0.650
8	-0.012	-0.021	3.3625	0.762
9	0.081	0.077	5.2575	0.631
10	0.100	0.104	8.1142	0.422
11	0.002	0.009	8.1150	0.520
12	0.034	0.031	8.4614	0.585
13	0.026	0.015	8.6400	0.655
14	0.048	0.040	9.3095	0.676
15	-0.068	-0.052	10.676	0.638
16	0.024	0.051	10.816	0.698
17	0.031	0.030	10.991	0.754
18	0.002	-0.000	10.992	0.811
19	0.048	0.054	11.555	0.830
20	-0.007	-0.021	11.574	0.864
21	0.009	0.012	11.595	0.898
22	-0.088	-0.033	14.040	0.828
23	-0.035	-0.042	14.411	0.851
24	0.014	0.008	14.474	0.884
25	-0.004	-0.010	14.478	0.912
26	0.032	0.024	14.795	0.927
27	0.004	-0.011	14.800	0.946
28	0.070	0.039	16.336	0.928
29	-0.017	-0.000	16.424	0.944
30	0.123	0.137	21.183	0.818
31	0.114	0.140	25.277	0.664
32	-0.068	-0.072	27.756	0.584
33	-0.033	-0.027	28.074	0.617
34	0.004	0.019	28.079	0.662
35	0.034	0.013	28.451	0.693
36	0.001	0.007	28.451	0.736

Como ya se sabe, para realizar estas pruebas estando en la ventana de la serie, en este caso U1, vamos a: View/Unit root test/, y seleccionamos tanto los test DF, ADF y PP. Tan sólo indicaremos que en virtud de que la serie de los residuos para esta regresión no presenta tendencia indicaremos para los tres test “None”, obteniendo los siguientes resultados:

- Para DF:

Cuadro 12.4.3.25.: Prueba de Raíz Unitaria Dickey-Fuller.

Augmented Dickey-Fuller Unit Root Test on U1

Null Hypothesis: U1 has a unit root
Exogenous: None
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-16.91242	0.0000
Test critical values:	1% level	2.573159	
	5% level	-1.941949	
	10% level	-1.615950	

*MacKinnon (1996) one-sided p-values.

- Para ADF:

Cuadro 12.4.3.26.: Prueba de Raíz Unitaria Augmented Dickey-Fuller

Augmented Dickey-Fuller Unit Root Test on U1

Null Hypothesis: U1 has a unit root
Exogenous: None
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-12.27466	0.0000
Test critical values:	1% level	-2.573188	
	5% level	-1.941953	
	10% level	-1.615948	

*MacKinnon (1996) one-sided p-values.

- Para PP:

Cuadro 12.4.3.27.: Prueba de Raíz Unitaria Phillips Perron.

Phillips-Perron Unit Root Test on U1

Null Hypothesis: U1 has a unit root
Exogenous: None
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-16.91238	0.0000
Test critical values:	1% level	-2.573159	
	5% level	-1.941949	
	10% level	-1.615950	

*MacKinnon (1996) one-sided p-values.

Conclusión: Como en las tres pruebas la probabilidad asociada a cada estadístico es menor a 0.05 y en virtud de que el valor de los estadísticos DF, ADF y PP son mayores a los valores críticos de MacKinnon al 5% concluimos que la serie de los residuos no presenta raíz unitaria, por tanto, es ruido blanco, es decir, tiene un comportamiento aleatorio.

Ahora realicemos el segundo modelo planteado siguiendo y/o aplicando los mismos pasos y criterios descritos en el punto IV para el primer modelo. Por razones de espacio, tan sólo mostraremos los resultados obtenidos dejando al lector la interpretación, misma que puede realizarse como práctica. De tal suerte que los resultados son los siguientes:

- Resultados de la ecuación de regresión:

Cuadro 12.4.3.28.: Resultados de la Regresión DLARA.

Dependent Variable: DLARA
Method: Least Squares
Date: 03/11/09 Time: 15:56
Sample(adjusted): 9 285
Included observations: 277 after adjusting endpoints
Convergence achieved after 18 iterations
Backcast: -29 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.005370	0.001333	-4.029194	0.0001
AR(7)	0.810983	0.039580	20.48957	0.0000
MA(7)	-0.949112	0.016044	-59.15614	0.0000
SMA(31)	0.127127	0.063141	2.013397	0.0451

R-squared	0.073251	Mean dependent var	-0.003994
Adjusted R-squared	0.063067	S.D. dependent var	0.038360
S.E. of regression	0.037130	Akaike info criterion	-3.734423
Sum squared resid	0.376377	Schwarz criterion	-3.682090
Log likelihood	521.2175	F-statistic	7.192729

Durbin-Watson stat	1.969314	Prob(F-statistic)		0.000116
Inverted AR Roots	.97	.61 -.76i	.61+.76i	-.22 -.95i
	-.22+.95i	-.87+.42i	-.87 -.42i	
Inverted MA Roots	.99	.93 -.09i	.93+.09i	.89+.28i
	.89 -.28i	.82 -.45i	.82+.45i	.71+.61i
	.71 -.61i	.62+.78i	.62 -.78i	.57 -.74i
	.57+.74i	.41+.84i	.41 -.84i	.23 -.91i
	.23+.91i	.05 -.93i	.05+.93i	-.14+.92i
	-.14 -.92i	-.22+.97i	-.22 -.97i	-.32+.88i
	-.32 -.88i	-.49+.79i	-.49 -.79i	-.64+.68i
	-.64 -.68i	-.77+.53i	-.77 -.53i	-.86+.37i
	-.86 -.37i	-.89 -.43i	-.89+.43i	-.92+.19i
	-.92 -.19i	-.94		

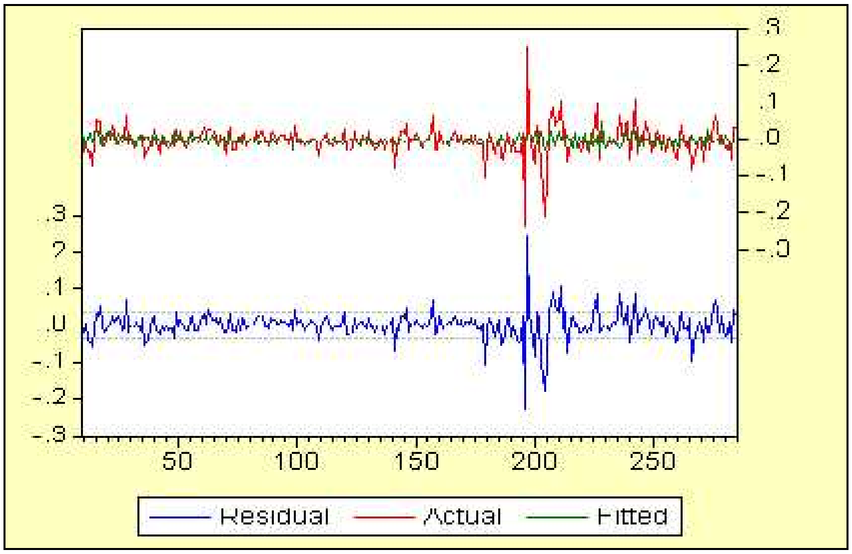
- <u>Matriz de covarianzas:</u>

Cuadro 12.4.3.29.: Matriz de Covarianzas

	C	AR(7)	MA(7)	SMA(31)
C	1.77660057177e-06	-6.74619695222e-06	9.79654783032e-07	-2.51274973952e-06
AR(7)	-6.74619695222e-06	0.00156660030031	-0.00029903445130	-0.000160631335892
MA(7)	9.79654783032e-07	-0.00029903445130	0.000257415768545	1.67250225397e-05
SMA(31)	-2.51274973952e-06	-0.00016063133589	1.67250225397e-05	0.00398672737268

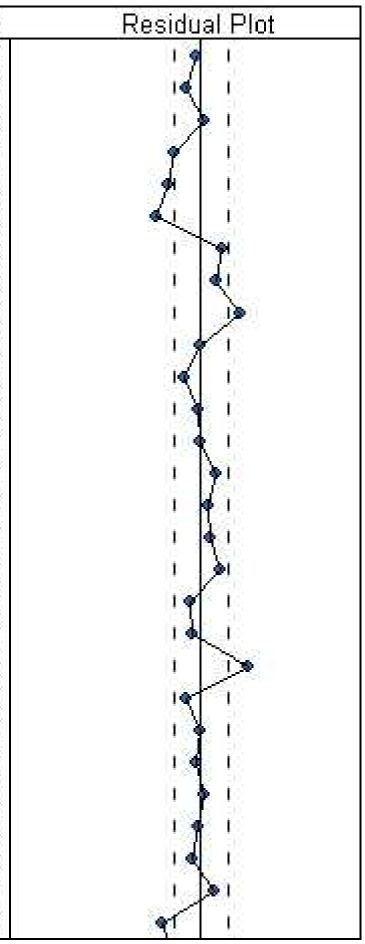
- <u>Tabla y gráfica de los residuos:</u>

Gráfica 12.4.3.10.: Residual, Actual y Ajuste de la variable.



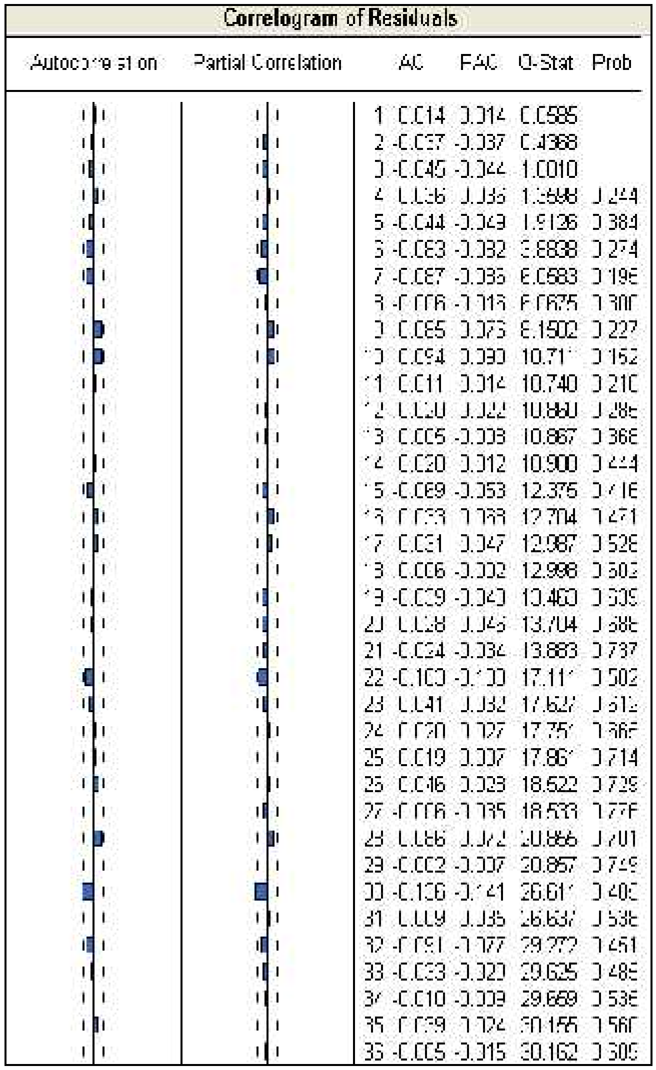
Gráfica 12.4.3.11.: Gráfica de parcela y residuales.

obs	Actual	Fitted	Residual
9	0.02036	0.02642	-0.00607
10	-0.03416	-0.01294	-0.02122
11	-0.01048	-0.01724	0.00676
12	-0.03575	-3.8E-05	-0.03571
13	-0.03049	0.01547	-0.04595
14	-0.07094	-0.01076	-0.06018
15	0.05486	0.02321	0.03166
16	0.04383	0.02177	0.02206
17	0.04635	-0.00735	0.05370
18	-0.01669	-0.01616	-0.00053
19	-0.01968	0.00407	-0.02375
20	0.01435	0.01845	-0.00410
21	0.00000	-0.00049	0.00049
22	0.03672	0.01333	0.02339
23	0.02375	0.01343	0.01032
24	-0.00252	-0.01509	0.01257
25	0.01253	-0.01361	0.02614
26	-0.00833	0.00571	-0.01404
27	0.00418	0.01447	-0.01029
28	0.06532	-0.00044	0.06576
29	-0.01257	0.00583	-0.01840
30	0.00787	0.00819	-0.00032
31	-0.02060	-0.01526	-0.00534
32	-0.00885	-0.01598	0.00713
33	0.00242	0.00397	-0.00155
34	-0.00081	0.01116	-0.01196
35	0.00723	-0.01098	0.01821
36	-0.05091	0.00291	-0.05382



- Correlograma de los residuos:

Cuadro 12.4.3.30.: Correlograma de residuales.



Correlogram of Residuals

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.014	0.014	0.0585	
		2	-0.037	-0.037	0.4368	
		3	-0.045	-0.044	1.0010	
		4	0.036	0.035	1.3698	0.244
		5	-0.044	-0.043	1.9126	0.384
		6	-0.083	-0.082	3.8838	0.274
		7	-0.087	-0.085	6.0583	0.195
		8	-0.006	-0.013	6.0675	0.300
		9	0.085	0.073	8.1502	0.227
		10	0.094	0.090	10.71	0.152
		11	0.011	0.014	10.740	0.210
		12	0.020	0.022	10.860	0.285
		13	0.005	-0.003	10.867	0.368
		14	0.020	0.012	10.900	0.444
		15	-0.069	-0.053	12.375	0.416
		16	0.033	0.033	12.704	0.471
		17	0.031	0.047	12.987	0.528
		18	0.006	-0.002	12.998	0.602
		19	-0.039	-0.040	13.460	0.639
		20	0.028	0.045	13.704	0.686
		21	-0.024	-0.034	13.883	0.737
		22	-0.100	-0.100	17.11	0.502
		23	0.041	0.032	17.627	0.612
		24	0.020	0.027	17.75	0.665
		25	0.019	0.007	17.86	0.714
		26	0.046	0.023	18.522	0.739
		27	-0.006	-0.035	18.533	0.776
		28	0.086	0.072	20.866	0.701
		29	-0.002	-0.007	20.867	0.749
		30	-0.136	-0.141	26.61	0.405
		31	0.009	0.035	26.637	0.536
		32	-0.091	-0.077	29.272	0.451
		33	-0.033	-0.020	29.625	0.485
		34	-0.010	-0.009	29.659	0.535
		35	0.039	0.024	30.156	0.560
		36	-0.005	-0.015	30.162	0.605

<u>Comentario sobre el correlograma:</u> Como puede observarse en las funciones de autocorrelación (AC) y autocorrelación parcial (PAC) ninguna sale de las bandas de confianza, pues en el rezago 30 ambas están en el límite; asimismo, los valores de AC y PAC son muy menores a 1; y la probabilidad asociada al estadístico Q son mayores que 0.05. Derivado de estos resultados podemos concluir que los residuos son ruido blanco.

- Test DF, ADF y PP para los residuos (en ese orden):

Cuadro 12.4.3.31.: Prueba de Raíz Unitaria Dickey-Fuller.

Augmented Dickey-Fuller Unit Root Test on U2

Null Hypothesis: U2 has a unit root
Exogenous: None
Lag Length: 0 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-16.47504	0.0000
Test critical values:	1% level	-2.573188	
	5% level	-1.941953	
	10% level	-1.615948	

*MacKinnon (1996) one-sided p-values.

Cuadro 12.4.3.32.: Prueba de Raíz Unitaria Augmented Dickey Fuller.

Augmented Dickey-Fuller Unit Root Test on U2

Null Hypothesis: U2 has a unit root
Exogenous: None
Lag Length: 1 (Fixed)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-12.16534	0.0000
Test critical values:	1% level	-2.573217	
	5% level	-1.941957	
	10% level	-1.615945	

*MacKinnon (1996) one-sided p-values.

Cuadro 12.4.3.33.: Prueba de Raíz Unitaria Phillips Perron.

Phillips-Perron Unit Root Test on U2

Null Hypothesis: U2 has a unit root
Exogenous: None
Bandwidth: 1 (Fixed using Bartlett kernel)

		Adj. t-Stat	Prob.*
Phillips-Perron test statistic		-16.47505	0.0000
Test critical values:	1% level	-2.573188	
	5% level	-1.941953	
	10% level	-1.615948	

*MacKinnon (1996) one-sided p-values.

X) Selección del modelo.

Ahora bien, si varios modelos pasan las pruebas de validación y se quiere discernir entre uno u otros, existen criterios para determinar cual de ellos es mejor. Es decir, los siguientes criterios se utilizan principalmente como elemento de comparación entre modelos alternativos para una misma variable, entre ellos tenemos:

4. Función logarítmica de verosimilitud (log likelihood):

Viene dado por:

$$L = \frac{-I}{2}\left[1 + \ln(2\text{®}) + \ln\frac{\Sigma e_i^2}{I}\right]$$

Donde:

I = Número de observaciones.

Σe_i^2 = Sumatoria de los errores de la regresión al cuadrado.

De acuerdo a Cabrer, *et al.-* (2001: 47) el valor que toma la función logarítmica de verosimilitud puede ser un criterio para valorar la bondad del modelo, esencialmente cuando se trata de comparar dos o más modelos alternativos. Pero, ¿Por qué no utilizar el R^2 para discernir entre un modelo u otro? Porque la principal ventaja de este estadístico frente al R^2 es que no se ve influido por las transformaciones a que se puede ver sometida la variable endógena. Por lo que se refiere a este criterio para la elección entre distintos modelos, se escoge aquel modelo que presenta un valor de la función logarítmica de verosimilitud (L) más alto.

5. <u>Criterio de información de Akaike (AIC: Akaike info criterion):</u>

Esta dado por:

$$AIC = \frac{2K}{I} - \frac{2L}{I}$$

Donde:

K= Número de variables exógenas incluidas en el modelo.

L = Valor de la función logarítmica de verosimilitud.

I = Número de observaciones.

El AIC sirve para comparar la bondad de ajuste entre dos modelos. El criterio para la elección entre distintos modelos se fundamenta en el valor estimado del estadístico. Se selecciona aquel modelo que presenta un estadístico AIC menor.

6. <u>Criterio de información Schwarz (Schwarz criterion):</u>

Esta dado por:

$$SC = \frac{K * \ln(I)}{I} - \frac{2L}{I}$$

Donde:

K= Número de variables exógenas incluidas en el modelo.

L = Valor de la función logarítmica de verosimilitud.

I = Número de observaciones.

Al igual que L y AIC, éste sirve para comparar la bondad de ajuste entre modelos alternativos. El criterio de elección de un modelo entre varios alternativos se fundamenta en el valor estimado del estadístico. Se selecciona aquel modelo que presenta un estadístico Schwarz menor.

También, para elegir entre un modelo y otro pueden considerarse la suma de cuadrados de los residuos (Sum squared resid) y el error estándar de la regresión (S. E. of regresión). Nos quedaremos con aquel modelo que presente la suma de los errores al cuadrado y el error estándar de la regresión menores. Así mismo, a pesar del problema de utilizar el R^2 como una prueba de la bondad de ajuste del modelo, dado que éste se ve fuertemente influenciado por las transformaciones de la serie, cuando se trata de discernir entre varios modelos nos quedaremos con aquel que posea el R^2 más alto.

De tal suerte, los resultados obtenidos en los dos modelos planteados son:

Cuadro 12.4.3.34.: Criterios para ARIMA y SARIMA.

Criterios	Primer modelo: ARIMA(7,1,7)	Segundo modelo: ARIMA(7,1,7)*SARIMA(0,0,31)
Suma de cuadrados de los residuos	0,383111	0,376377
Desviación estándar de la regresión	0,037393	0,03713
Coeficiente de determinación	0,056671	0,073251
Coeficiente de determinación ajustado	0,049785	0,063067
Logaritmo de verosimilitud (L)	518,7615	521,2175
Criterio de información de Akaike (AIC)	-3,72391	-3,734423
Criterio de Schwarz (SC)	-3,684661	-3,68209

En virtud de estos resultados y de acuerdo a lo establecido previamente para la selección de un modelo, no cabe duda de que el mejor modelo para estimar el precio de las acciones de Consorcio ARA es el segundo, a saber: ARIMA(7,1,7)*SARIMA(0,0,31). Pues, como puede observarse en el cuadro anterior: la suma de cuadrados de los residuos en el segundo modelo es menor que en el primero; la desviación estándar de la regresión también es menor en el segundo; el segundo modelo presenta un R^2 y $\overline{R}^2$ mayor que el primero; asimismo el segundo modelo presenta un mayor valor de L; el segundo presenta un menor valor de AIC; y en el caso del SC el primer modelo presenta un valor menor que el segundo. Dado que el segundo modelo cumple con la mayor parte de los criterios establecidos, no cabe duda sobre que modelo utilizar.

Estructura del modelo

Antes de seguir adelante es preciso establecer la estructura del modelo. En nuestro caso, el modelo tiene la siguiente estructura:

DLARA = -0.005370 + 0.810983*DLARA(-7) + Residuo$_t$ – (-0.949112)*Residuo(-7) – 0.127127*Residuo(-31) - (-0.949112)*(0.127127*Residuo(-38))

DLARA$_t$ = -0.005370 + 0.810983*DLARA(-7) + Residuo$_t$ + 0.949112*Residuo(-7) – 0.127127*Residuo(-31) + 0.949112*(0.127127*Residuo(-38))

La interpretación de la ecuación anterior es que el precio de las acciones de Consorcio ARA se explica en función de si misma 7 días antes (hábiles o de operación del mercado de valores mexicano), así como por el error que comete el modelo 7 y 31 días anteriores.

Y con esta ecuación de regresión es con la que vamos a proceder a predecir el precio de las acciones de ARA.

XI) Predicción.

Una vez seleccionado y planteado el modelo definitivo, es decir, aquel que cumple satisfactoriamente los criterios de evaluación establecidos, puede pasarse a la etapa de predicción.

La predicción puede realizarse en dos formas diferentes:

1.- De forma estática o paso a paso; y

2.- De forma dinámica o en cadena.

La diferencia entre una y otra es que la primera nos permite predecir sólo un periodo por delante y utiliza los valores reales para estimar y predecir, por ejemplo, para nuestro caso, tan sólo nos proporcionará el precio de las acciones para la observación 286; mientras que la segunda permite al modelo que vaya realimentando sus propias predicciones, es decir, permite predecir para varias observaciones de una sola vez. Ambas predicciones (estática y dinámica) coinciden para el primer periodo, pero, a partir del segundo, la predicción dinámica utiliza el valor estimado y no el valor real del periodo precedente.

Para nuestro caso, tal y como hemos venido trabajando o más bien dicho, generando la serie, Eviews no la reconoce como transformación y diferenciación de la serie ordinaria (original) ARA, es decir, por ejemplo, a la serie DLARA la reconoce como una serie ordinaria (original) y no como una transformación de la serie ARA. Por lo que, al realizar la predicción, Eviews tan sólo nos permitirá realizarla para la serie DLARA y no para la serie ARA. Por ejemplo, al realizar la predicción de la serie DLARA tal y como hasta el momento hemos venido trabajando, Eviews nos desplegará el siguiente cuadro de diálogo:

Cuadro 12.4.3.35.: Especificación de estimación.

Como puede observarse, Eviews nos desplegará una ventana para la predicción de DLARA teniendo nosotros que realizar las transformaciones necesarias para obtener los valores de la predicción de la serie ARA.

Por tanto, en virtud de que Eviews no reconoce la serie DLARA como una transformación y diferenciación de la serie ordinaria (original) ARA hay que generar una nueva serie. Así, en el menú principal vamos a: Quick/Generate Series…/Ok, y en la venta que despliega escribimos:

dlogara=dlog(ara)

En donde, nuevamene, estamos generando la variable DLARA, pero en términos de que Eviews la reconozca como una serie transformada de la serie original ARA. Entonces, la indicación “dlog(ara)” está generando la primera diferencia de la serie en logaritmos del precio de las acciones de ARA. En este contexto, Eviews reconocerá a la serie ARA como una serie ordinaria y la serie DLOG(ARA) la reconocerá como una transformación y diferenciación de la serie ordinaria ARA. Con ello, cuando realicemos la predicción Eviews nos proporcionará dos opciones de predicción, a saber: 1) En términos de la serie ordinaria, es decir, ARA y por tanto en pesos por acción; y 2) En términos de la serie transformada y diferenciada, es decir, DLOG(ARA) teniendo nosotros que realizar los cálculos necesarios que nos permitan obtener el precio de las acciones de ARA. Nótese que siempre, de aquí en adelante, vamos a trabajar con la serie DLOG(ARA) y no con DLOGARA. Sin embargo, ésta última es la que aparecerá en el Workfile. Pero no nos preocupemos, ya que Eviews reconoce ésta serie aunque no aparezca en el Workfile. Además, siempre podemos trabajar en la barra de comandos.

Así, una vez generada la nueva serie [DLOG(ARA)], misma que por cierto deberá ser exactamente igual a la serie DLARA como puede observarse en el siguiente cuadro:

Cuadro 12.4.3.36.: Series

Group: UNTITLED Workfil...

View | Procs | Objects | Print | Name | Freeze | Transform

obs	DLARA	DLOGARA
1	NA	NA
2	0.017094	0.017094
3	-0.024883	-0.024883
4	-0.015762	-0.015762
5	-0.032290	-0.032290
6	0.026092	0.026092
7	0.040893	0.040893
8	-0.005128	-0.005128
9	0.020357	0.020357
10	-0.034162	-0.034162
11	-0.010480	-0.010480
12	-0.035750	-0.035750
13	-0.030487	-0.030487
14	-0.070938	-0.070938
15	0.054862	0.054862
16	0.043830	0.043830
17	0.046354	0.046354
18	-0.016689	-0.016689
19	-0.019679	-0.019679
20		

En el menú principal vamos a: Quick/Estimate Equation…/Ok., y escribimos:

dlog(ara)_c_ar(7)_ma(7)_sma(31)/Ok.

Donde "_" significa espacio.

Obteniendo el siguiente cuadro de resultados:

Cuadro 12.4.3.37.: Análisis de Regresión del Modelo AR(7) MA(7) SMA(31).

Dependent Variable: DLOG(ARA)
Method: Least Squares
Date: 03/16/09 Time: 18:58
Sample(adjusted): 9 285
Included observations: 277 after adjusting endpoints
Convergence achieved after 18 iterations
Backcast: -29 8

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.005370	0.001333	-4.029194	0.0001
AR(7)	0.810983	0.039580	20.48957	0.0000
MA(7)	-0.949112	0.016044	-59.15614	0.0000
SMA(31)	0.127127	0.063141	2.013397	0.0451

R-squared	0.073251	Mean dependent var	-0.003994
Adjusted R-squared	0.063067	S.D. dependent var	0.038360
S.E. of regression	0.037130	Akaike info criterion	-3.734423
Sum squared resid	0.376377	Schwarz criterion	-3.682090
Log likelihood	521.2175	F-statistic	7.192729
Durbin-Watson stat	1.969314	Prob(F-statistic)	0.000116

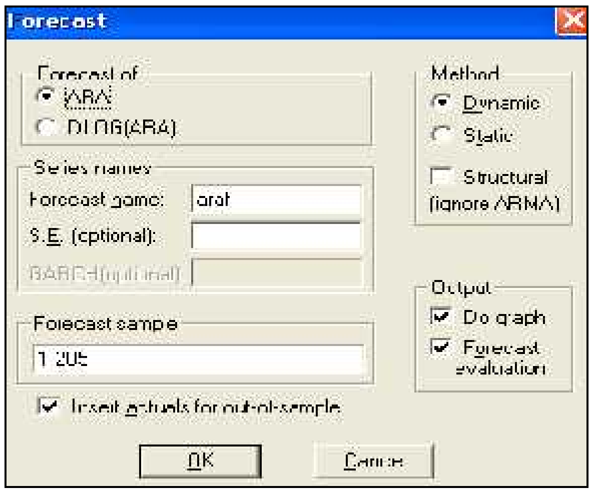
Inverted AR Roots	.97	.61 -.76i	.61+.76i	-.22 -.95i
	-.22+.95i	-.87+.42i	-.87 -.42i	
Inverted MA Roots	.99	.93 -.09i	.93+.09i	.89+.28i
	.89 -.28i	.82 -.45i	.82+.45i	.71+.61i
	.71 -.61i	.62+.78i	.62 -.78i	.57 -.74i
	.57+.74i	.41+.84i	.41 -.84i	.23 -.91i
	.23+.91i	.05 -.93i	.05+.93i	-.14+.92i
	-.14 -.92i	-.22+.97i	-.22 -.97i	-.32+.88i
	-.32 -.88i	-.49+.79i	-.49 -.79i	-.64+.68i
	-.64 -.68i	-.77+.53i	-.77 -.53i	-.86+.37i
	-.86 -.37i	-.89 -.43i	-.89+.43i	-.92+.19i
	-.92 -.19i	-.94		

El mismo resultado puede obtenerse desde la barra de comandos, en la cual escribiremos:

ls_dlog(ara)_c_ar(7)_ma(7)_sma(31)/Ok.

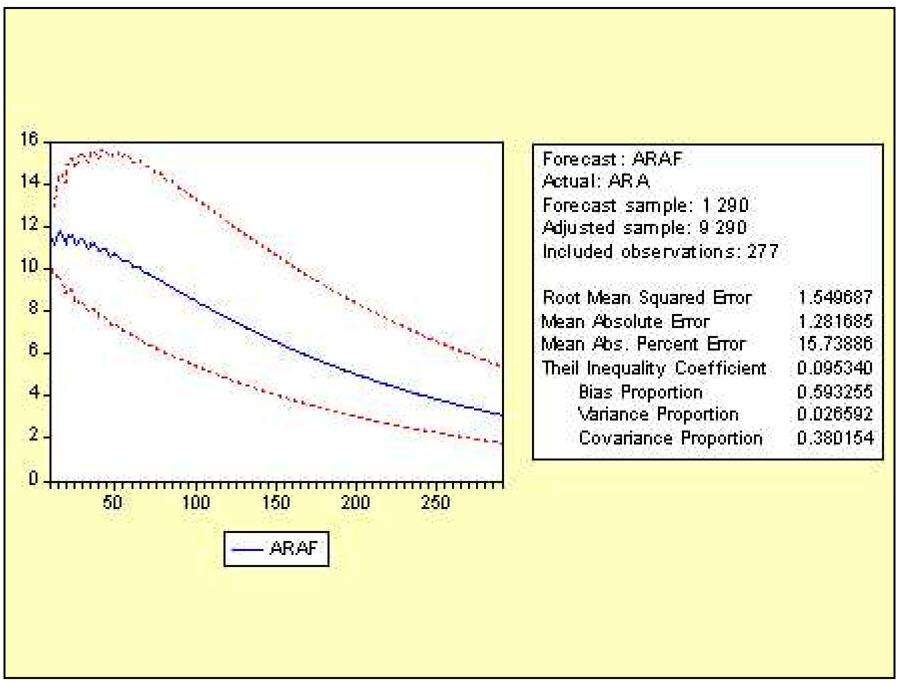
En donde: “ls” indica estimación por mínimos cuadrados y “_” espacio.

Cuadro 12.4.3.37.: Especificación de estimación.



En la ventana de la ecuación presionamos el botón “FORECAST”, lo cual desplegará una ventana como la anterior. Como puede observarse en esta ventana a diferencia de la primera (ver página 38), Eviews nos da ha escoger entre dos términos de predicción (Forecast of), a saber: 1) En términos de la serie ARA; y 2) En términos de la serie transformada y diferenciada de ARA [DLOG(ARA)], en nuestro caso seleccionaremos ARA. Luego, nos solicita el nombre de la serie que predecirá (Series names), el cual por default Eviews nos da: araf, donde la “f” indica la predicción, es decir, forecast. En seguida, el programa nos solicita el tamaño de la muestra para la predicción (Forecast sample); el cual para nuestros fines modificaremos de 1 a 290. Y, también nos solicita el método de predicción (Method), es decir, si la predicción será dinámica (Dynamic) o estática (Static); para nuestro caso, en virtud de que nos interesa predecir para la observación 286 a 290, seleccionaremos el método dinámico. El resto lo dejamos tal cual lo específica el programa por default/Ok. La nueva variable aparecerá en el Workfile y asimismo, nos desplegará una gráfica que muestra estadísticos que permiten determinar la bondad de la predicción. Así, la gráfica es:

Gráfica 12.4.3.12.: Gráfico de estimación.



Como puede observarse, la serie ARAF está dentro de las bandas de confianza por lo que podemos concluir que la predicción es buena. Asimismo, con respecto a las estadísticas, dado que casi todas (con la excepción de root mean squared error, mean absolute error y mean abs, percent error) tienen valores menores a uno, decimos que es buena estimación de la predicción.

El siguiente cuadro muestra los precios reales de las acciones de ARA para la observación 281 a 290, la estimación y predicción realizada con nuestro modelo con el método dinámico para las mismas observaciones, así como el error de predicción:

Cuadro 12.4.3.38.: Serie de datos reales.

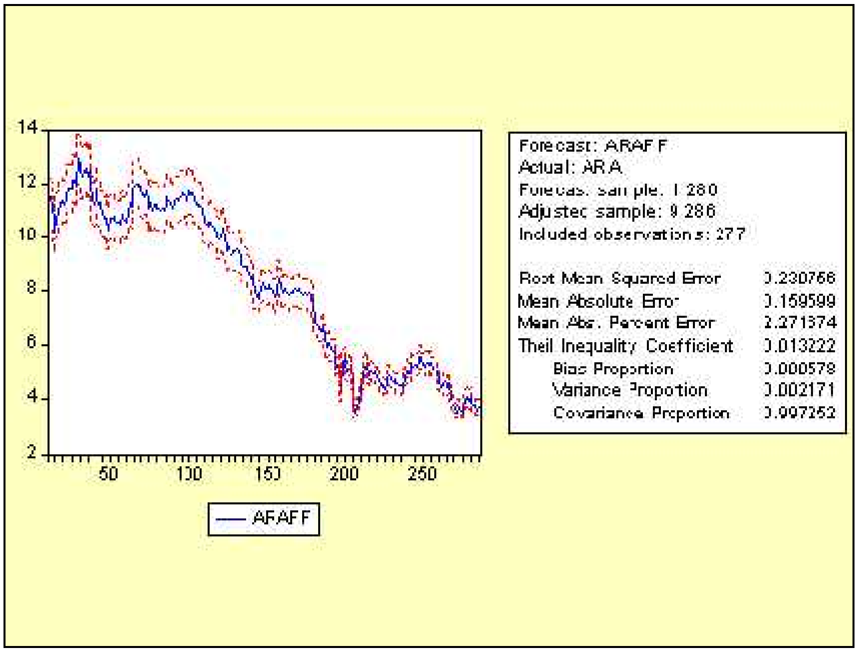
Obs.	Fecha	Precio real (Pr)	Precio estimado (Pe)	Error (Pr-Pe)
281	13/Feb/2009	3.87	3.25	**0.62**
282	16/Feb/2009	3.86	3.23	**0.63**
283	17/Feb/2009	3.64	3.21	**0.43**
284	18/Feb/2009	3.76	3.20	**0.56**
285	19/Feb/2009	3.86	3.18	**0.68**
286	20/Feb/2009	3.71	3.16	**0.55**
287	23/Feb/2009	3.67	3.15	**0.52**
288	24/Feb/2009	3.75	3.13	**0.62**
289	25/Feb/2009	3.65	3.11	**0.54**
290	**26/Feb/2009**	**3.83**	**3.09**	**0.74**

Observese que el error de estimación (para el caso de las observaciones 281 a 285) y el error de predicción (para el caso de las observaciones 286 a 290) es grande, en promedio de 0.589 pesos por acción, con el método dinámico. Así, la predicción nos proporciona una guía para determinar, en este caso, la compra o venta de acciones de Consorcio ARA. Téngase presente, además, que los modelos ideales para predecir series financieras son los ARCH y GARCH. Sin embargo, puede utilizarse la metodología Box & Jenkins para este tipo de series recomendándose la predicción para tres días posteriores y no más (Reyes, 2009).

Ahora bien, podemos realizar la predicción con el método estático con la restricción de que tan sólo nos proporcionará la estimación para la siguiente observación, en nuestro caso, la 286. Por lo que, si queremos estimar para la predicción 287 tenemos que ingresar el precio real de las acciones de ARA para la observación 286, es decir, necesitamos alimentar constantemente la serie. Sin embargo, este método es más preciso en virtud de que utiliza el valor real anterior en lugar de utilizar el valor o los valores estimados como el método dinámico.

Para llevar a cabo la predicción con el método estático estimamos la ecuación y oprimimos el botón “Forecast” en donde únicamente modificaremos el nombre de la serie por “ARAFF”; el método por el de estático (Static); y la muestra para la predicción por 1 a 286 (aunque el lector puede poner de 1 a 290 para que observe como tan sólo el programa predecirá para la observación 286)/Ok. Lo cual nos desplegará una gráfica como la siguiente y en el Workfile aparecerá la serie “araff”:

Gráfica 12.4.3.13.: Gráfico de proyección.



Note usted en la gráfica el precio estimado de las acciones de ARA (araff) con el método estático ésta más pegado a las bandas de confianza y ello se debe a que a través de este método de estimación, estático, se utilizan los valores reales y no los estimados. Por lo que la estimación con este método será más precisa. Asimismo, con respecto a las estadísticas, dado que casi todas (con la excepción de mean abs, percent error) tienen valores menores a uno, decimos que es buena estimación de la predicción.

El siguiente cuadro muestra el precio real, estimado y de predicción con el método estático para las observaciones 281 a la 286:

Cuadro 12.4.3.39.: Serie de datos reales.

Obs.	Fecha	Precio real (Pr)	Precio estimado (Pe)	Error (Pr-Pe)
281	13/Feb/2009	3.87	3.98	**-0.11**
282	16/Feb/2009	3.86	3.83	**0.03**
283	17/Feb/2009	3.64	3.81	**-0.17**

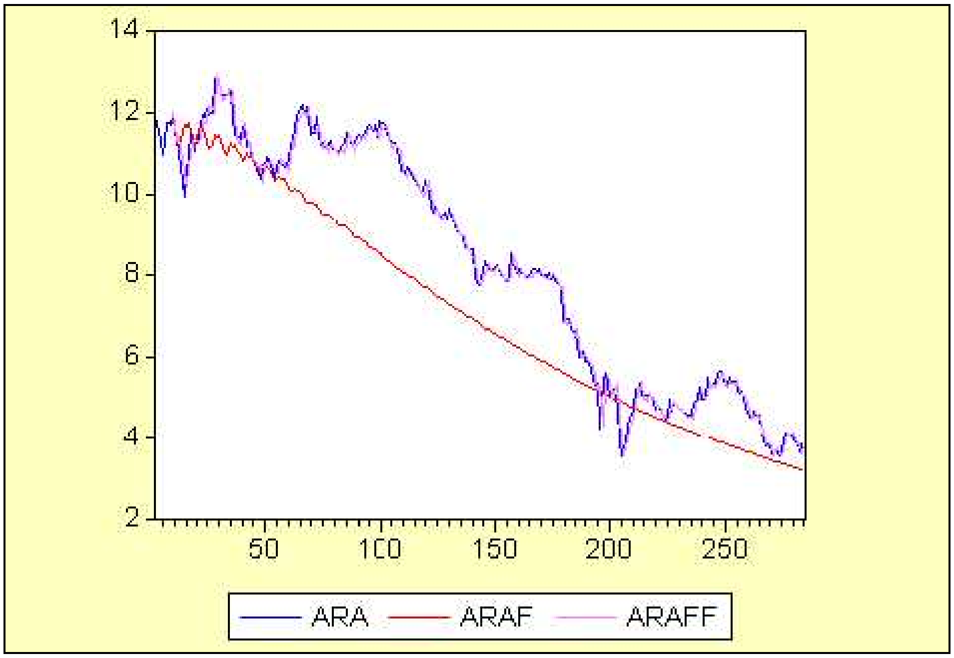
284	18/Feb/2009	3.76	3.61	**0.15**
285	19/Feb/2009	3.86	3.77	**0.09**
286	**20/Feb/2009**	**3.71**	**3.79**	**-0.08**

Como se desprende del siguiente cuadro, el error de estimación (para el caso de las observaciones 281 a 285) y el error de predicción (para el caso de la observación 286) es muy pequeño. Por ejemplo, para el caso de la predicción del precio de las acciones del 20 de febrero de 2009 el error entre el precio real y el de predicción es de -0.08 pesos por acción, es decir, 8 centavos de peso. Error muy inferior al error del precio de predicción para la misma fecha con el método dinámico, a saber, 55 centavos de peso por acción.

En síntesis, para el caso de la predicción para la observación 286 (20 de febrero de 2009) el mejor método, es decir, el que se acerco más al precio real de las acciones de Consorcio ARA es el estático. Sin embargo, el método dinámico nos proporciona una guía sobre el precio futuro de las acciones y por tanto, decidir comprar o vender estas acciones.

Finalmente, el siguiente gráfico nos proporciona el precio real de las acciones de Consorcio ARA (línea de color azul), el precio estimado con el método dinámico (ARAF, línea de color rojo) y con el método estático (ARAFF, línea de color rosa), mismo que se obtiene seleccionando en el Workfile dichas series y dando click derecho sobre las mismas: Open/as Group/Ok., lo cual nos desplegará un cuadro con los valores de las series. Una vez ahí, vamos a: View/Graph/Line/Ok., operación que desplegará:

Cuadro 12.4.3.14.: Comportamiento gráfico de las variables “ARA”, “ARAF” y “ARAFF”.



Como puede observarse la predicción con el método estático (ARAFF) es más exacta que la predicción con el método dinámico (ARAF), ello debido a que, como se comento, el primero utiliza los valores reales para la estimación posterior y el segundo utiliza los valores estimados para las predicciones posteriores.

APENDICE A: HERRAMIENTAS MATEMATICAS BASICAS

1 Conjuntos y números reales.

***Definición de conjunto**:* de manera sencilla diremos que un conjunto es un grupo de objetos; digamos, podemos decir que los números pares entre 3 y 15 es un conjunto, los cuales son 4,6,8,10,12,14, donde a cada uno de estos números se le llama miembro o elemento del conjunto.

Un conjunto se especifica listando sus miembros, en cualquier orden, dentro de llaves. El conjunto anterior se pone así: {4,6,8,10,12,14}, mismo que podemos identificar con la letra A. Así, decimos por ejemplo que un conjunto A es un subconjunto de un conjunto B $\Leftrightarrow$ todos los elementos de A también son elementos de B. Si A= {6,8,10} y B = {4,6,8,10,12,14}, entonces A es un subconjunto de B.

Ahora bien, ciertos conjuntos de números tienen nombres especiales. Los números 1,2,3,4,5, ...,etc. forman el **conjunto de los enteros positivos o números naturales =** {1,2,3,...}.

Los puntos suspensivos indican que la lista de elementos no tiene fin, aun cuando sabemos cuáles son los elementos del conjunto. Por otra parte, los enteros positivos junto con el cero y los enteros negativos –1,-2,-3,-4.... forman el conjunto de los enteros ={……-4,-3,-2,1,0,1,2,3,4,....}.

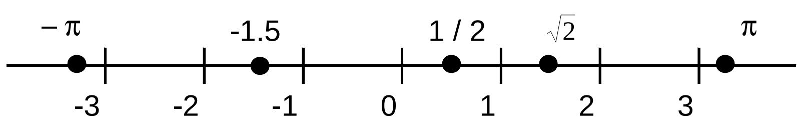
Igualmente, el conjunto de los *números racionales,* consiste en números como 1/2 y 1/8, que se pueden escribir como una razón (cociente) de dos enteros, es decir, un número racional es aquel que puede escribirse como p/q, donde p y q son enteros y $q \neq 0$ ya que no se puede dividir entre cero. Los números 17/19, -3/8 y –5/-3 son racionales; el entero 3 es racional ya que 3 = 3/1. Todos los enteros son racionales.

Los números racionales se pueden representar mediante números **decimales conmensurables** (con un número definido de cifras), tales como 3/4= 0.75 y 3/2

= 1.5 o mediante **decimales inconmensurables periódicos** (con un grupo de dígitos que se repiten indefinidamente), tales como 2/3= 0.66666...., -4/11 = 0.3636....., y digamos 2/15 = 0.13333.....

También, los números que se representan mediante **decimales inconmensurables no periódicos se llaman números irracionales.** Un número irracional no se puede escribir como un entero dividido entre otro entero. Los números π= 3.1416 y $\sqrt{2}$ son irracionales.

Los dos**, tanto los números racionales como los números irracionales forman el conjunto de los números reales.** Estos números pueden representarse mediante **puntos en una recta y se les llaman coordenadas**. Ejemplo:



II.2 Ecuaciones.

Significado: una ecuación es un planteamiento que señala que *dos expresiones son iguales:* Cada una de las *expresiones se llama lado o miembr*o y están separadas por el signo de igualdad (=).

Ejemplos de ecuaciones:

a) x+2=3 ; b) y/y-5=7 ; c) w=7-z ; d) x^2 + 3x+2= 0

Vemos que cada ecuación contiene cuando menos una **variable**, donde ésta se representa generalmente (pero no necesariamente) por las letras finales del alfabeto latino y se **define** como aquella literal que puede tomar cualquier valor dentro de un dominio o rango que se específica previamente.

Decimos que las ecuaciones a) y b) son ecuaciones en las variables ***x*** y ***y,*** respectivamente. La ecuación c) se da en las variables ***w*** y ***z:*** En la ecuación a) x+2=3, se les llama **constantes** a los números 2 y 3 por que su valor es fijo.

Es importante decir que una ecuación *siempre debe estar definida;* es decir, nunca se permite que una variable tenga un valor para el cual cualquier *expresión* de la ecuación resulta indefinida. Así, en el caso de la ecuación y/(y-5)=7, la ***y*** no puede ser 5 por que provocaría que el denominador fuera 0.

Por otra parte, se *entiende* **por resolver una ecuación,** el encontrar todos los valores de sus variables para los cuales la ecuación se verifica; a estos valores los llamamos **soluciones de la ecuación** y decimos que la satisfacen.

Cuando una letra representa un número o cantidad *desconocida* en una ecuación se le denomina **incógnita.** Así, en las ecuaciones a) x+2=3, la variable x es la incógnita; sólo el valor 1 para x satisface la ecuación. A esta solución se le llama **raíz** y cómo es una sola solución escribimos $[1]$. En c) w=7-z es una ecuación con dos incógnitas: ***w*** y ***z***. Una solución es el par de valores: w=4 y z= 3; no obstante, existe una cantidad infinita de soluciones; igualmente, en d) $X^2+3x+2=0$, la raíz o solución es –2 porque al sustituir ***x*** por -2 se verifica la ecuación: $(-2)^2+3(-2)+2=0$.

Derivado de lo anterior (con base en las soluciones encontradas) también podemos decir que una ecuación es un conjunto de restricciones impuestas a cualquiera de las variables que la integran.

Finalmente, es conveniente decir que una ecuación debe resolverse para las incógnitas; dichas soluciones son los valores encontrados para las variables que, al sustituirlos en lugar de la incógnita, satisfacen la ecuación.

Ecuación lineal.

Una ecuación lineal en la variable ***x*** puede escribirse así:

$$Ax+b=0$$

Donde a y b son constantes y $a \neq 0$.

También se le denomina de *primer grado por que el exponente o mayor potencia de la variable x es uno.*

Ecuación cuadrática.

*Una ecuación cuadrática en la variable **x** podemos escribirla de la siguiente manera:*

$$ax^2 + bx + c = 0.$$

Donde a,b y c son constantes y $a \neq 0$.

Es de segundo grado dado que el exponente máximo de la variable **x** es dos**.**

En general podemos decir que, por ejemplo:

$3x + 4 = 0$, es una ecuación lineal de primer grado;

$X^2 + X + 12 = 0$, es una ecuación cuadrática de segundo grado cuya curva se llama parábola.

$X^3 - X = 0$, es una ecuación cúbica de tercer grado; y

$X^n = 0$, es una ecuación de grado e-nésimo.

II.2. 1. Solución de ecuaciones.

Dado que por definición todas las ecuaciones expresan una *igualdad* del lado izquierdo con el lado derecho, todas las operaciones en uno de los lados, también deben realizarse del otro lado para que se mantenga la *igualdad.* Cualquier número puede sumarse o sustraerse de un lado de la ecuación, siempre y cuando el mismo número sea agregado o restado del otro lado.

Simultáneamente ambos lados pueden multiplicarse o dividirse por el mismo número, ser elevados a la misma potencia o sacarles la raíz cuadrada a ambos lados. Cuando se resuelven las ecuaciones se debe tener cuidado con la remoción de los paréntesis y de las fracciones.

Solución de ecuaciones lineales:

a) Si tenemos: $3\text{-}5(2X - 1)(X - 3) = 43 - 10X^2$

Resolvemos: $3 - 5(2X^2 - 5X - 3) = 43 - 10X^2$

$3 - 10X^2 + 25X + 15 = 43 - 10X^2$

$25X = 25$

$X = 1$

b) Sea $5X - 6 = 3X$, si sumamos $-3X$ en los dos lados:

$5X - 6 + (-3X) = 3X + (-3X)$

$2X - 6 = 0$

$X = 3$

c) Digamos que $2(p + 4) = 7p + 2$

$2p + 8 = 7p + 2$

$2p = 7p + 2 – 8$

$-7p + 2p = -6$

$-5p = - 6$

$p = -6/-5$

$p = 6/5$

d) Si decimos que la ecuación $S = P + Prt$ es la fórmula para el valor S de una inversión de un capital P a una tasa de interés anual simple (r) durante un periodo de (t) años, entonces, si queremos conocer el capital, despejamos P así:

$$S = P + Prt$$

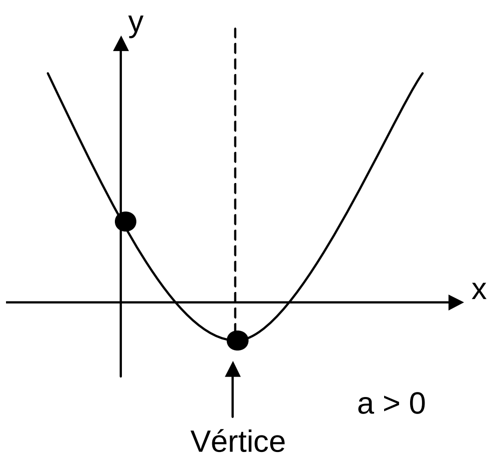
$$S = P\ (1 + rt)$$

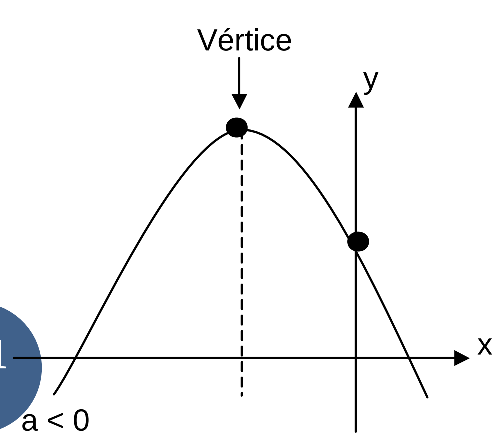
$$P = S\ /\ (1 + rt)$$

Solución de una ecuación cuadrática* [3]*:

La gráfica o curva de la función cuadrática $Y = f\ (X) = ax^2 + bx + c = 0$ se denomina parábola, misma que tiene las siguientes características:

1. Cuando a es mayor que 0, la curva abre hacia arriba;
2. Cuando a es menor que 0, la curva abre hacia abajo;
3. El vértice es $(-b/2a,\ f\ (-b\ /\ 2a))$; y
4. La ordenada en el origen (intercepción) es **c.**





El punto en que la parábola $Y = ax^2 + bx + c$ intercepta el eje **Y,** es el de la ordenada en el origen y se obtiene cuando damos a x el valor de 0. O sea que las coordenadas de la intercepción con el eje **Y,** son (0, c).

Así, si la ecuación cuadrática es $Y = f(x) = ax^2 + bx + c$ sustituyendo en, $-x^2 - 4x + 12$, vemos que a = -1, b = -4 y c = 12. Puesto que a es menor que 0, la parábola abre hacia abajo. La coordenada X del vértice es $-b / 2a = -(-4) / 2(-1) = -2$. La coordenada **Y** es $f(-2) = -(-2)^2 - 4(-2) + 12 = 16$. Así, el vértice o punto más alto de la parábola es (-2, 16). Como c = 12, la ordenada en el origen (x= 0) es 12.

Para determinar las abscisas en el origen, se iguala **Y** a 0, es decir $Y = -x^2 - 4x + 12$ y con ello determinamos el valor de **x:**

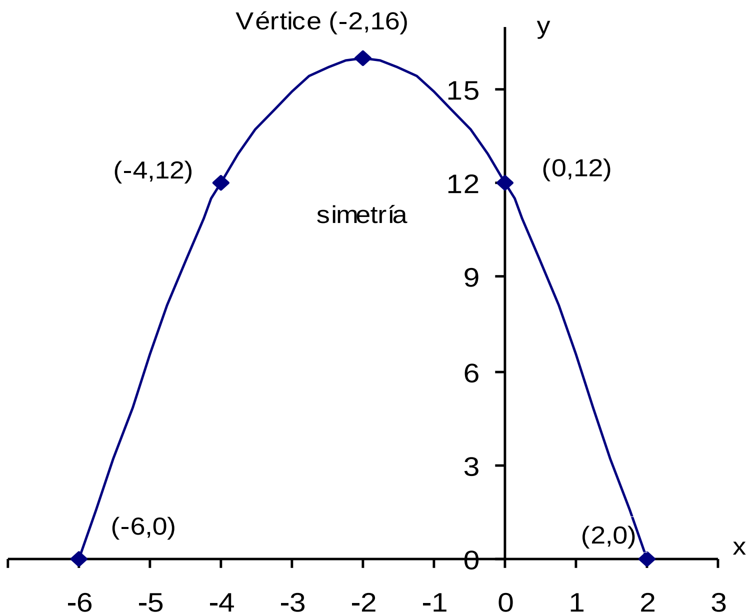
$$0 = -x^2 - 4x + 12$$

$$0 = -(x^2 + 4x - 12)$$

$$0 = -(x + 6)(x - 2)$$

Luego entonces $X_1 = -6$;   también, $X_2 = 2$

Así tenemos el vértice en (-2, 16), las intercepciones en (-6, 0) y (2, 0). Como el punto de intercepción al eje **Y** es (0, 12) denominado *ordenada en el origen,* está a dos unidades a la derecha del vértice, (-2, 16), existe un punto a la izquierda del mismo con igual ordenada (-4, 12), con el que se obtiene la ***simetría con respecto al vértice.***



Solución de ecuaciones cuadráticas usando la fórmula cuadrática.

En la ecuación $ax^2 + bx + c = 0$, donde a, b y c son constantes y a diferente de cero, entonces la fórmula es:

$$\mathbf{X = -b} \pm \sqrt{b^2 - 4ac} \ \mathbf{/ \ 2a;}$$

a) Resolver $4x^2 - 17x + 15 = 0$ aquí tenemos: a = 4; b = -17 y c = 15

Luego: $X = -(-17) \pm \sqrt{(-17)^2 - 4(4)(15)} \ / 2(4) = 17 \pm \sqrt{49} \ / 8$

Las raíces son: $X_1 = 17+7 / 8 = 24 / 8 = 3$ y $X_2 = 17 - 7 / 8 = 10/8 = 5/4$

b) $3x^2 - 5x = 2$
$3x^2 - 5x - 2 = 0$ a = 3, b = -5 y c = -2

Así: $X = 5 \pm \sqrt{25+24} / 6 = 5 \pm \sqrt{49} / 6 = 5 + 7 / 6$

Luego $X_1 = 5+7/6 = 2$ y $X_2 = 5-7/2 = -2/6 = -1/3$

Sistema de ecuaciones o ecuaciones simultáneas.

Cuando una situación se describe en forma matemática, en ocasiones se expresa por medio de un conjunto de ecuaciones. Al respecto, es importante decir que un conjunto de ecuaciones simultáneas se puede resolver normalmente con álgebra ***si el número de ecuaciones es igual al número de variables.*** Así para resolver dos variables se requieren dos ecuaciones; generalizando, resolver para n variables, significa que se requieren n ecuaciones. Sin embargo, en ocasiones aun cuando el número de ecuaciones es igual al número de variables no se encuentra su solución. Lo anterior se ilustra con los siguientes ejemplos:

Sistema de Ecuaciones Lineales:

i) ***Inconsistente***:

x-y = *2............. (1)*
x-y = 5.............. (2)
El sistema es inconsistente porque *es igual a 2 o 5.*

ii) ***Dependiente***:

$2x + 3y = 6$...... (1)
$6x + 9y = 18$..... (2)
Hay una relación o dependencia de 1 a 3 de la ecuación uno a la dos, o viceversa.

iii) ***Independientes y consistentes:***

$x + y = 5$................ (1)
$x – y = 1$............... (2)

Solo tienen solución los sistemas de ecuaciones que son ambas independientes y consistentes, cuya solución puede ser por los métodos de: **1. Eliminación; 2. Sustitución y 3. Determinantes.**

1. *Por eliminación de una variable por adición o sustitución:*

$5x + 2y = -5$........... (1)
$-3x + 4y = 29$................ (2)

Si multiplicamos (1) por 2:

$10x + 4y = -10$............. (1) por 2
$-3x + 4y = 29$.................. (2)

Solución $13x + 0 = -39$

$x = -3$

Comprobación:

Sustituimos (- 3) por X en (1) y tenemos

$5(-3) + 2y = -5$
$-15 + 2y = -5$
$y = 5$

Respuesta $X = -3$; $Y = 5$

2. Por sustitución de una variable:

Ejemplo: $3x – 4y = 5$................. (1)
$X + 7y = 10$................ (2)

De la segunda ecuación x = 10 – 7y, luego sustituyendo (10 – 7y) por x en (1) tenemos:

$$3(10-7y) - 4y = 5 \ldots\ldots\ldots\ldots (1)$$
$$30 - 21y - 4y = 5$$
$$30 - 25y = 5$$
$$30 - 5 = 25y$$
$$25 = 25y$$
$$y = 1$$

Sustituimos 1 por y en (1):

$$3x - 4(1) = 5$$
$$x = 3$$

Respuesta x = 3 ; y = 1

3. Por determinantes:

Referencias: Sabemos que una ***matriz*** se define como un arreglo rectangular de números y que una matriz cuadrada es aquella en que el número de “renglones” es igual al número de “columnas”: A cada matriz cuadrada **A corresponde un sólo valor conocido como su determinante** $|A|$ **.**

3.1. Tipos de determinantes:

Determinante de segundo orden, se define como:

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

Vemos que el valor de $|A|$ está dado por la diferencia de dos productos cruzados. Con números:

$$\begin{vmatrix} 4 & 5 \\ 1 & 3 \end{vmatrix} = 4(3) - 5(1) = 7$$

Determinante de tercer orden, se define en términos de determinante de segundo orden:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$|A| = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

Ejemplo:

$$|A| = \begin{vmatrix} 1 & 5 & 3 \\ 2 & 0 & 5 \\ -4 & 1 & -2 \end{vmatrix} = 1\begin{vmatrix} 0 & 5 \\ 1 & -2 \end{vmatrix} - 5\begin{vmatrix} 2 & 5 \\ -4 & -2 \end{vmatrix} + 3\begin{vmatrix} 2 & 0 \\ -4 & 1 \end{vmatrix}$$

$|A|$ = (0-5) – 5 (-4+20) + 3 (2 – 0) = -5+20 - 100 + 6 - 0 = -105 + 26

$|A|$ = - 79

En general decimos que un determinante de orden **n** se define en términos del determinante de orden (n-1).

Ejemplos: Resuelva el siguiente sistema de ecuaciones simultáneas para **X** e **Y**.

3x –4y = 5

x +7y = 10

X = $\begin{vmatrix} 5 & -4 \\ 10 & 7 \end{vmatrix}$ entre $\begin{vmatrix} 3 & -4 \\ 1 & 7 \end{vmatrix}$ = 35 – (-40)/ 21 – (-4) = 75 / 25 = 3

X = 3

Para Y = $\begin{vmatrix} 3 & 5 \\ 1 & 10 \end{vmatrix}$ entre $\begin{vmatrix} 3 & -14 \\ 1 & 7 \end{vmatrix}$ = 30 – 5 / 21- (-4) = 25/ 25

Y = 1

Regla de Cramer.

Al método de resolver un par de ecuaciones lineales de dos incógnitas se puede extender al caso general de una ecuación con **n** incógnitas. Si Ax = h

La solución es $X_1 = |A1|$ entre $|A|$

Donde $|A|$ es el *determinante base o determinante de la matriz coeficiente* $|A|$, y $|A1|$ representa el determinante de la matriz coeficiente A con el coeficiente de X_1 reemplazado por la columna de las constantes h_1, h_2, h_3........, h_n.

$X_2 = |A2|$ entre $|A|$ y en general $X_j = |Aj|$ entre $|A|$

Observe que cuando la matriz A es singular, es decir, $|A| = 0$ no se puede usar la **Regla de Cramer**.

Ejemplo: Encuentre los valores de las tres incógnitas (X, Y, Z) de las siguientes ecuaciones usando la Regla de Cramer:

x + y +z = 6

5x –y +27 = 9

3z +6y –5z = 0

Los ponemos en notación matricial:

$$\begin{vmatrix} 1 & 1 & 1 \\ 5 & -1 & 2 \\ 3 & 6 & -5 \end{vmatrix} \begin{vmatrix} x \\ y \\ z \end{vmatrix} = \begin{vmatrix} 6 \\ 9 \\ 0 \end{vmatrix}$$

$$X = \begin{vmatrix} 6 & 1 & 1 \\ 9 & -1 & 2 \\ 0 & 6 & -5 \end{vmatrix} \text{ entre } \begin{vmatrix} 1 & 1 & 1 \\ 5 & -1 & 2 \\ 3 & 6 & -5 \end{vmatrix} = 57/57 = 1$$

$$Y = \begin{vmatrix} 1 & 6 & 1 \\ 5 & 9 & 2 \\ 3 & 0 & -5 \end{vmatrix} \text{ entre } \begin{vmatrix} 1 & 1 & 1 \\ 5 & -1 & 2 \\ 3 & 6 & -5 \end{vmatrix} = 114/57 = 2$$

$$Z = \begin{vmatrix} 1 & 1 & 6 \\ 5 & -1 & 9 \\ 3 & 6 & 0 \end{vmatrix} \text{ entre } \begin{vmatrix} 1 & 1 & 1 \\ 5 & -1 & 2 \\ 3 & 6 & -5 \end{vmatrix} = 171/57 = 3$$

Sistema no lineal, también llamado solución de ecuaciones simultáneas cuando al menos una de las ecuaciones es cuadrática.

Un sistema de ecuaciones simultáneas de un tipo más complejo se puede resolver usando el *método de sustitución*, ejemplo:

$$X^2 + 4y = 20.............. (1)$$

$$2x = 8............ (2)$$

Sustituya (8+2x) por **y** en (2) en la primera ecuación:

$$X^2 + 4(8+2x) = 20........... (1)$$

$$X^2 + 32 + 8x = 20$$

$$X^2 + 8x +12 = 0$$

$$(X + 6) (X + 2) = 0$$

Obtenemos ya sea x = -6 o x= -2

Si x = -6, la segunda ecuación se convierte en:

$$Y -2(-6) = 8.......... (2)$$

$$Y +12 = 8$$

$$Y = -4$$

Cuando x = -2, la segunda ecuación se convierte en:

$$Y -2(-2) = 8............. (2)$$

$$Y + 4 = 8$$

$$Y = 4$$

Respuesta: hay dos soluciones: ya sea x = -6 con y = -4 o también: x = -2, y = 4.

Ejemplo 2: resolver el sistema $X^2 -2x + y -7 = 0........ (1)$

$$3x – y + 1 = 0........... (2)$$

Despejando “**y”** en la segunda ecuación obtenemos:

$$y = 3x + 1 \ldots\ldots\ldots\ldots\ldots\ldots\ldots (3)$$

Sustituyendo en la (1) y simplificando:

$$X^2 - 2x + (3x + 1) - 7 = 0$$

$$X^2 + x - 6 = 0$$

$$(x + 3)(x - 2) = 0$$

Se obtiene $X = -3$ o bien $X = 2$

De la ecuación (3), cuando $x = -3$, tenemos $y = -8$; también, cuando $x = 2$, $y = 7$.

II.3 El Concepto de una Función [2 y 3].

Como señalan E. Haeussler y R.S. Paul, el matemático G.W. Leibniz en el siglo XVII introdujo el concepto de Función dentro del vocabulario matemático; una **función** es un tipo especial de relación de entrada y salida, o dicho en otras palabras: insumo y producto.

Definición: una función es una **regla** que asigna a cada número de entrada exactamente un número de salida. El conjunto de todos los números de entrada a los cuales se aplica la regla se le llama dominio de la función. Al conjunto de todos los números de salida se le llama ámbito o *contradominio*.

Ejemplo: Al invertir dinero a una tasa de interés, el interés I, (salida), depende del tiempo t (entrada) en que se invierte el dinero. Para expresar esta dependencia decimos: "I es función de t". Así, supongamos que $100.00 producen interés simple a una tasa anual del 6%; escribimos:

$$I = 100(i)\, t$$

$$I = 100(0.06)\, t \ldots\ldots\ldots\ldots (1)$$

Donde I se expresa en pesos y t en años. Si $t = 1/2$, obtenemos:

$$I = 100(0.06)1/2 \ldots\ldots\ldots\ldots\ldots (2)$$

En este caso, en (2) se asigna a la entrada t el valor ½ y a la salida 3. En (1) se define la **regla**: el multiplicar t por 100(0.06). Dicha en otra forma $t \rightarrow I$, también: $t \rightarrow 100(0.06)t$.

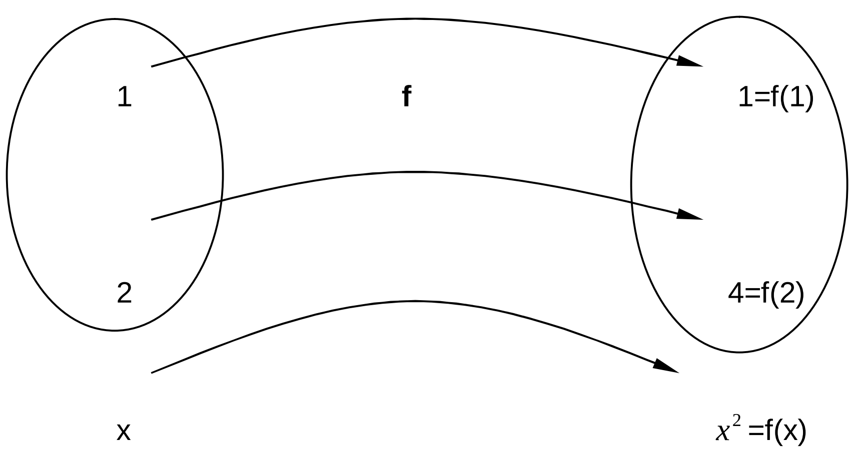
De lo anterior observamos que el número de entrada t no puede ser negativo porque (-) no tiene sentido. Luego entonces el dominio consiste en todos los números no negativos, t ≥ 0. En la ecuación (2), cuando la entrada es ½, la salida es 3, es decir, es la parte del ámbito o *contra dominio.*

A la variable t que representa números de entrada para una función se le llama **variable independiente** y a la variable I que representa números de salida se le denomina **dependiente**, pues su valor depende del valor de la variable independiente; decimos que la segunda es función de la primera, en otras palabras, la salida es *función* de la entrada.

Por costumbre se usa la letra ***f*** para representar reglas de funciones, pero si lo deseamos, podemos usar otras letras. En el caso anterior escribimos I=f(t).

Debe comentarse que f(t) no significa f multiplicado por t, sino que f(t) significa la salida que corresponde a t.

Lo anterior lo podemos generalizar diciendo que gráficamente lo expresamos de la siguiente manera: digamos si f(x)=X^2 y si recordamos que una función es una correspondencia mediante la cual se asigna exactamente un número de salida en el contradominio a solo uno de los números de la entrada del dominio, tenemos:



II.3:1. Funciones especiales.

Función constante:

Una función constante de la forma k(x) = c, en donde c es una constante, se denomina *función constante.*

Ejemplo: j(x)=3, el dominio de j son todos los números reales.

Como dicen E.F. Haeussler y R. S. Paul[3]: "las funciones constantes pertenecen a una clase más amplia de funciones", denominadas funciones *polinomiales.* En general decimos que una función de la siguiente forma:

$$F(x)=C_nX^n + C_{n-1}X^{n-1}+........+C_1X^1+ C_0$$

Donde n es un número entero no negativo y Cn, C_{n-1},......, C_0 son constantes con $Cn \neq 0$ se llaman función polinomial en (x). A ***n*** se le llama ***grado* de la función** y a cada Ci, *coeficiente.* Luego $f(x) = 4X^2 - 6X + 3$ es una función polinomial de segundo grado con coeficiente inicial 4;

También: H(x) = 2-3X es una función de primer grado. En este caso se dice que la primera es una *función cuadrática* y la segunda, *función lineal.*

Con base en lo anterior es que podemos decir que en una ecuación se establece una relación particular entre dos o más expresiones algebraicas. Las ecuaciones pueden contener cualquier número de variables. Con la terminología matemática se expresa como:

$$Y=f(X).$$

Se dice que Y es función de X, es decir el valor de Y depende del valor que tome X; cuando su valor depende de más de una variable escribimos:

$$Y = f(x, z, h).$$

En este caso el valor de Y depende del valor que tomen las variables (x, z, h).

Como puede observarse, lo anterior puede formalizarse y generalizarse para más variables que llamaremos explicativas o independientes.

II.3.2. Funciones exponenciales y logarítmicas.

II.3.2.1. Funciones exponenciales.

Estas se caracterizan por contener una constante (b) elevada a un exponente variable (x), es decir, $f(x) = b^x$, donde "b" es mayor que 0 y también "b" es $\neq 1$, "x" es cualquier número real; también decimos que cuando es mayor que 1, la curva $Y=b^x$ asciende de izquierda a derecha, en otras palabras, al aumentar "**x"** también lo hace "**y". ** Cuando $0 \leq b \leq 1$, entonces, en ese caso la curva de $Y=b^x$ desciende de izquierda a derecha, se observa un comportamiento distinto de la variable "**y"** al del caso anterior, es decir, al aumentar "x", el valor de la variable "y" disminuye y toma valores cercanos a 0.

Un ejemplo típico del uso de una función exponencial es su aplicación a los intereses que gana un capital invertido durante cierto número de años a una tasa de interés determinada. Hablamos de **interés compuesto** [(3)] cuando: "el interés que percibe una suma de dinero invertida (capital o monto especial) se reinvierte, de manera que este interés también gana interés". Es decir, el interés *se compone* o convierte en capital y, por ello, “hay interés sobre interés”.

Por ejemplo supóngase que se invierten $100 dólares (o cualquier otra unidad monetaria) a la tasa del 5% compuesto anualmente. Al final del primer año, el valor de la inversión es el capital original ($100) más el interés generado por este:

$$[100(0.05)] = 100 + 100(0.05) = \$105.00$$

Esta es la cantidad sobre la que se genera interés para el segundo año. Al final del segundo periodo anual, el valor de la inversión es el capital que se tenía al final del primer año, $105 más el interés producido por esa cantidad:

$$[105(0.05)] = 105+105(0.05) = \$110.25.$$

Así, el capital se incrementa en 5% cada año. Los $110.25 representan el capital original, más todo el **interés acumulado; se le denomina monto acumulado o monto compuesto.** La diferencia entre el monto compuesto y el capital original se denomina ***interés compuesto .***Así, el interés compuesto aquí es:

$$110.25 – 100.00= 10.25.$$

En términos más generales, si se invierte un capital “P” a una tasa de 100r por ciento compuesto anualmente (por ejemplo, al 5%, r es 0.05), el monto compuesto después de un año será:

$$P + Pr \text{ o bien } P (1 + r).$$

Al final del segundo año, el monto compuesto es:

$$P\ (1+r) + P\left[(1+r)\right] = P\ (1+r) + \left[1+r\right] = P\ (1+r)^2$$

Esta operación continúa. Después de tres años, el monto compuesto es:

$$P\ (1+r)^3.$$

En general, el ***monto compuesto*** "S" de un capital "P" al final de **"n"** años, a la tasa de "**r"** compuesta anualmente, está dado por:

$$S = P\ (1+ r)^2 \ldots\ldots\ldots (1)$$

Por considerar importante distinguir las ecuaciones de interés compuesto de las de **interés simple**, diremos que [2] calculamos éste último, cuando "el dinero es invertido a interés simple, los pagos de interés son hechos ***cada año sobre la inversión original***"***.*** Por consiguiente si "P" pesos son invertidos a una tasa de "**r"** por "**t**" años, el valor de la inversión, "**A"**, al final del periodo, será:

$$A = P + Prt$$

Ejemplo: Calcule el valor de una inversión de $500.00 después 3 años si el interés simple se paga anualmente a una tasa anual del 5%.

$$A = P + Prt$$

$$A = 500 + (500*5/100*3)$$

$$A = 575 \text{ pesos}$$

II.4 Funciones Explícitas e Implícitas.

Cualquier función que es de la forma ***Y=f(X)*** es conocido como una función explícita de Y. Claramente se ve que el valor de Y depende del valor de X, y decimos que mientras que X es una variable independiente, Y es una variable dependiente.

En ciertos casos no se distingue con claridad la relación funcional entre las variables dependiente e independiente. Por ejemplo, $2x^2+3xy+y^2=0$, es una función de la forma f(x, y)=0 conociéndose como una función implícita.

En ocasiones una función explícita puede, pero no siempre, derivarse de funciones implícitas simples, por ejemplo, la función implícita $2x^2 + 3y - 8 = 0$, produce dos funciones explícitas:

$$Y = \frac{8-2X^2}{3} \quad \text{y} \quad X = \sqrt{4-\frac{3Y}{2}}$$

II.5 El Rango o dominio de una Variable.

Para muchas funciones de la forma ***Y = f(X)*** la variable independiente X puede tener cualquier valor negativo, cero o positivo. En tales casos decimos que el rango o dominio de la variable X es el *conjunto de todos los números enteros*. Ejemplos de este tipo de funciones son:

$$Y = 3x$$

$$Y = 2x^2 - 5$$

$$Y = x^3 - 3x^2 + 3x - 1$$

En otros casos deseamos limitar el rango de la variable independiente. Por ejemplo la función $Y=\sqrt{X}$.

Si Y es un número real X debe tener un valor positivo. En este caso X se limita al rango de números positivos. El rango o dominio está determinado cuando X>0.

Si el movimiento de la variable independiente se restringe a cierto rango, entonces se deben fijar los límites superiores e inferiores del rango.

Así, si nos interesan todos los valores de x entre x=a, y x=b, el rango es $a<x<b$ o hablando más estrictamente $a \leq x \leq b$ donde x puede ser igual a cualquiera de los límites o éstos contener a ellos.

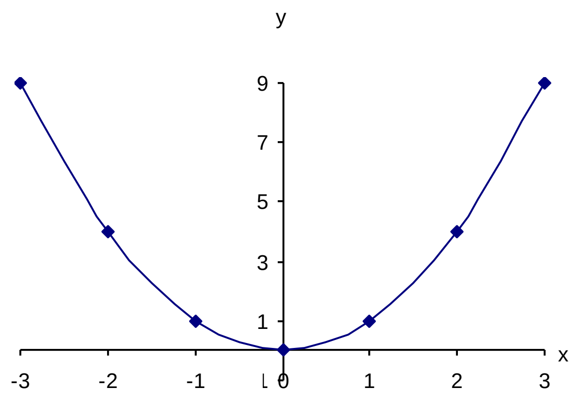
En funciones algebraicas se supone que la variable es continua cuando puede dividirse o fraccionarse, y es discreta cuando no se puede dividir o fraccionar.

II. 6 Simetría[3].

La simetría forma parte del análisis sobre el comportamiento gráfico de las ecuaciones.

Así por ejemplo:

1. Si tenemos la ecuación $Y= X^2$ y si le damos valores a X de –1,0 y 1, obtenemos para Y los valores 1,0 y 1, respectivamente. Si graficamos estos valores observamos que la parte o porción que está del lado izquierdo del eje de las "y" es el espejo o imagen sobre el eje "y" de la parte que se halla al lado derecho del mismo, y viceversa. Con literales diremos que si (X_0, Y_0) es cualquier punto en la curva, entonces el punto $(-X_0, Y_0)$ también está en la curva, por lo que decimos que es una **curva simétrica con respecto al eje "y".**

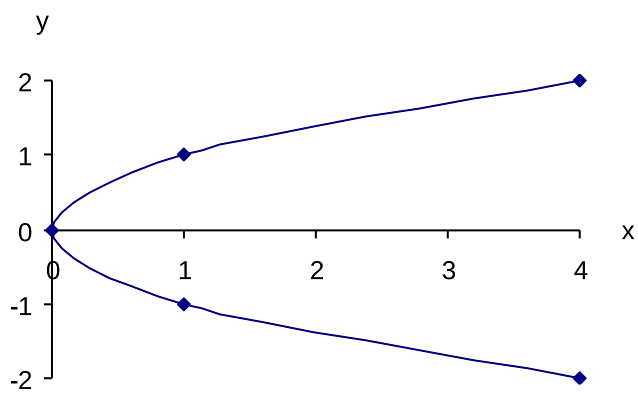


$y = x^2$ $y = x^2$

Definición: Una curva es simétrica con respecto al eje "y" $\Leftrightarrow$ $(-X_0,Y_0)$ queda en la curva cuando (X_0,Y_0) también queda.

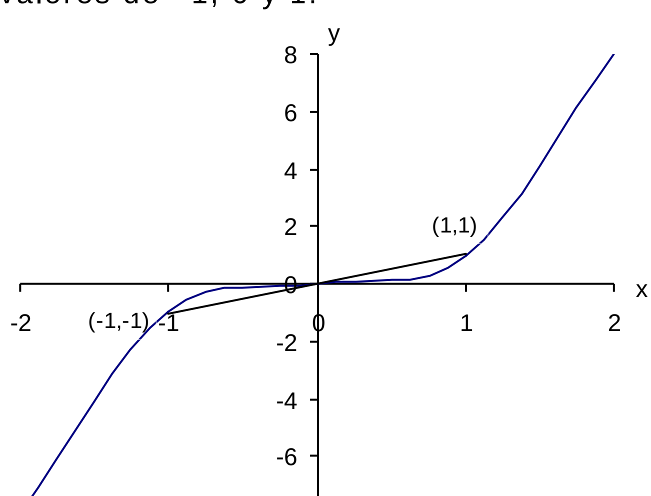
2. De manera complementaria podemos decir que si tenemos la ecuación $X=Y^2$ podemos probar la simetría con respecto al eje "x", dando valores a Y, digamos: -1,0 y 1, en cuyo caso X toma los valores 1,0 y 1, que al graficarlos obtenemos una curva en la que vemos que a cada punto en ella, con coordenadas $(X_0,-Y_0)$ corresponde otro punto en la misma curva con coordenadas (X_0,Y_0).

$x = y^2$



Luego entonces decimos que una curva es simétrica con respecto al eje de las "x" $\Leftrightarrow$ $(X_0, -Y_0)$ está en la curva cuando (X_0, Y_0) también lo está.

3. Cuando tenemos $Y=X^3$ y le damos valores a X de –1,0 y 1, obtenemos para Y los valores de –1, 0 y 1:



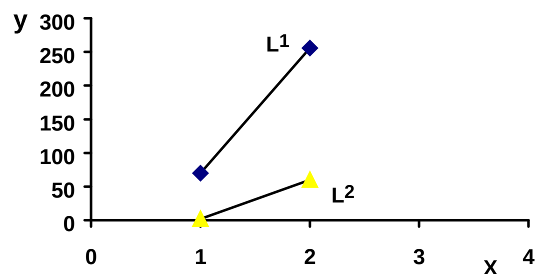
En este caso se dice que la curva es simétrica con respecto al “origen” porque si el punto (X, Y) está en la curva, vemos que el punto con coordenadas (-X, -Y) también lo está; en consecuencia, el segmento de recta que une los puntos (X, Y) y (-X,-Y) es bisecado por el origen.

Definición: una curva es simétrica con respecto al “origen” $\Leftrightarrow$ el punto (-X,-Y) está en la curva cuando el punto (X, Y) también lo está.

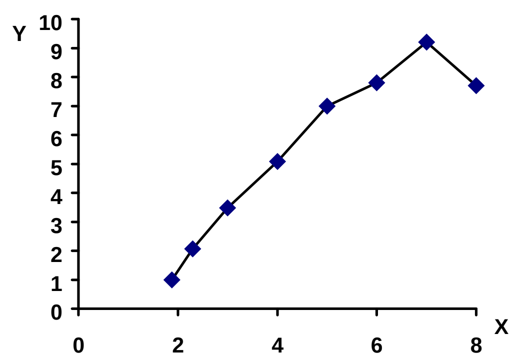
II.7 Rectas, Parábolas y Sistemas[3].

II.7.1 Rectas.

Se pueden representar en forma conveniente muchas relaciones entre cantidades mediante rectas. Una característica de una recta es su inclinación. Por ejemplo, en el siguiente diagrama la recta L_1 tiene una mayor inclinación que la recta L_2, ello significa que L_1 tiene mayor inclinación con respecto al eje horizontal.



Para medir la inclinación de una recta se utiliza el concepto de **Pendiente.** Así, si tenemos los puntos (2, 1) y (4, 5) podemos trazar la recta L.



Se observa que la coordenada x aumenta de 2 a 4 y la coordenada Y aumento 1 a 5. La tasa promedio de variación de Y con respecto a x es la razón:

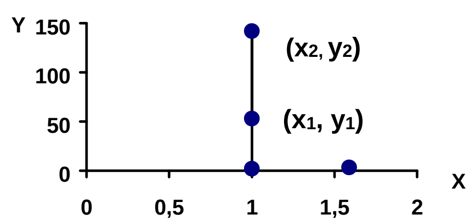
$$\frac{Cambio \to en \to Y}{Cambio \to en \to X} = \frac{cambio \to vertical}{cambio \to horizontal} = \frac{5-1}{4-2} = \frac{4}{2} = 2$$

Ello indica que para cada aumento unitario en X se tiene un aumento de 2 unidades en Y. Por ello la recta asciende de izquierda a derecha. Se dice que la pendiente de la recta es 2.

Definición: Sean (X_1 y Y_1) y (X_2 y Y_2) dos puntos sobre una recta en donde $X_1 = X_2$, la pendiente de la recta es el número dado por:

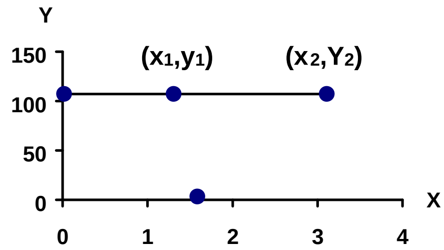
$$m = \frac{Y_2 - Y_1}{X_2 - X_1} = \frac{cambio\,vertical}{cambio\,horizontal} \quad \ldots\ldots\ldots\ldots. (1)$$

No se define la pendiente de una recta vertical, pues $X_1 = X_2$ tal que el denominador en (1) es cero.



Donde $X_1 = X_2$

Para una recta horizontal, dos puntos cualquiera tiene $Y_1 = Y_2$, por lo que el numerador en (1) es cero; y decimos que m = 0



$Y_1 = Y_2$

Con base en lo antes expuesto podemos resumir diciendo que:

Formas de ecuaciones de rectas.

Forma punto-pendiente	$Y-Y_1= m(X-X_1)$
Forma pendiente-intercepción "Y"	$Y= mX+b$
Forma lineal general	$Ax+By+C=0$
Recta vertical	$X=0$
Recta horizontal	$Y=b$

II.7.1.1 Ejercicios.

1. Encontrar la ecuación de la recta que tiene pendiente 2 y pasa por (1,-3).

   Sabemos que m=2 y que $(X_1, Y_1) = (1,-3)$, utilizando la forma punto-pendiente, escribimos $Y-(-3)=2(X-1)$, resolviendo tenemos:

Y+3=2X-2

Luego llegamos a la forma lineal general 2X-Y-5=0.

2. Hallar la ecuación de la recta que pasa por los puntos (-3,8) y (4,-2).

   Aquí tenemos que m=-2-8/4-(-3)=-10/7

   Si tomamos las coordenadas del primer punto (-3.8) como (X_1, Y_1) en la forma punto-pendiente, escribimos Y-8=-10/7 $\left[X-(-3)\right]$, haciendo operaciones tendremos lo siguiente:

   Y-8=-10/7(X+3)

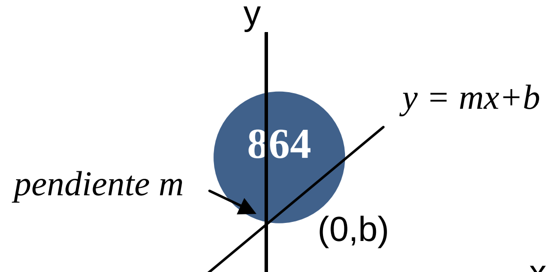
   7Y-56=-10X-30

   O bien:

   10X+7Y+26=0

   Ahora, si escogemos las coordenadas del segundo punto (4,-2) como (X_1, Y_1) también llegamos al mismo resultado.

3. Cuando tenemos el punto (0, b) decimos que una recta corta el eje “Y” y se denomina **intercepción Y;** asimismo indicamos que b es la ordenada en el origen.



Si conocemos la pendiente (m) y la ordenada en el origen (b) de una recta, la ecuación es: Y -b=m(X-0).

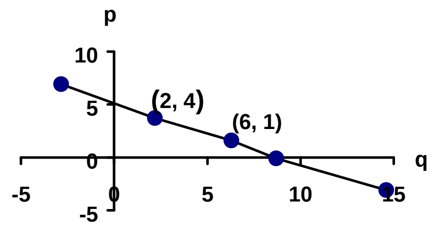
Si despejamos Y tenemos Y = mX+b, llamada forma pendiente-intercepción "y".

Ejemplo económico:

Sea X = q = cantidad; luego si (2, 4) = (q_1, p_1)

Y = p = precio (6, 1) = (q_2, p_2)

Tenemos que $m=\frac{1-4}{6-2}=\frac{-3}{4}=-\frac{3}{4}$



Ello significa que por cada aumento unitario en q, ocasiona una disminución de ¾ en P.

En resumen podemos decir que una recta:

1. Tiene pendiente cero, cuando la recta es horizontal;
2. Tiene pendiente indefinida, cuando la recta es vertical;
3. Tiene pendiente positiva, cuando la recta asciende de izquierda a derecha.
4. Tiene pendiente negativa, cuando la recta desciende de izquierda a derecha.

La ecuación de una recta con pendiente 3 y ordenada en el origen -4 es:

$$Y = mx + b$$

$$Y = 3x + (-4)$$

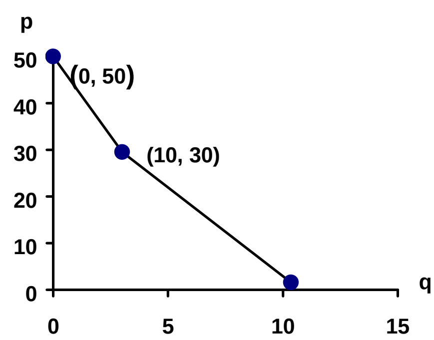
$$Y = 3x - 4$$

Aplicaciones:

Un fabricante dispone de 100 kilos de material con el que puede producir dos productos, A y B, que requieren de 4 y 2 kilos de material por unidad, respectivamente. Si X y Y denotan A y B, entonces todos los niveles de producción están dados por los cambios de X y Y que satisfacen la ecuación:

$$4x + 2y = 100 \text{ en donde } x,\ y \geq 0.$$

Despejando Y, se obtiene la forma pendiente-ordenada en el origen Y = -2x + 50 donde m=-2 indicando la tasa de variación del nivel de producción de B con respecto al nivel de producción de A, así si fabricamos una unidad más de A, se requieren 4 kilos más de material, lo que daría por resultado 4/2=2 unidades menos de B. Por consiguiente, al aumentar X en una unidad, Y disminuye en 2 unidades. Para trazar su gráfica de Y=-2x+50 puede utilizar la intercepción y(0, 50) y si cuando x=10, y=30, tenemos:



Uso de la sumatoria $\sum$

Como se usará mucho, sabemos que con propiedad se escribe $\sum_{i=1}^{n}$ donde i=1,2,...n, y que n = tamaño de la muestra; sin embargo, en lo sucesivo y por motivos prácticos simplemente se escribirá $\sum$, sin que se olvide su representatividad apropiada.

APENDICE B: FUNDAMENTOS DE PROBABILIDAD

Contexto e importancia

¿Por qué estudiar la teoría de la probabilidad?

Porque nos permite *predecir* (Richmond et al, 1964: 101) con cierto grado de confianza los resultados posibles de un experimento; también porque nos ayuda en la toma de decisiones inteligentes en la economía y en los negocios, cuando existen el riesgo y la incertidumbre (Salvatore, 1990). En este contexto, al seleccionar probabilísticamente una muestra (n) de una población (N), con sus datos podemos *estimar* las características de esta última con un error de muestreo determinado a priori. Esto significa predecir de antemano y de acuerdo con el método de muestreo usado para seleccionarla (Sánchez, 2005).

Por consiguiente, continuando con este razonamiento, Lind et al (2005:140) indica que la probabilidad es el cálculo de que un evento (resultado posible) *ocurra en el futuro,* cálculo que se hace a partir de una muestra (n) proveniente de una población (N), y que por eso al procedimiento utilizado se le denomina estadística inferencial. Conjugándolos, podemos decir que es conveniente usar esta última en situaciones en que hay incertidumbre, dado que se mide con los métodos de la teoría probabilística.

Puesto que en la toma de decisiones siempre hay incertidumbre, es importante evaluar científicamente todos los riesgos involucrados. En esta evaluación resulta útil la teoría de la probabilidad también conocida como teoría de la incertidumbre.

Dicho en otras palabras, es muy importante su participación en la economía al brindarnos la oportunidad de que al trabajar con una muestra se puedan estimar las características de la población con cierta probabilidad, y que por ese motivo a la probabilidad se le haga sinónimo de factibilidad, posibilidad o confianza en los resultados obtenidos con una porción de la población o universo estadístico.

La teoría de la probabilidad proporciona la herramienta necesaria para sustentar las inferencias estadísticas que se hagan a partir de una muestra. Podemos usarla para explicar el comportamiento de un *evento o suceso* que tiene incertidumbre. El evento tiene incertidumbre cuando es estocástico o aleatorio, cuando se gesta en un experimento que produce *eventos o resultados* que son aleatorios porque son “unos de tantos resultados posibles”, cuyo total de resultados constituye el “espacio muestral”. Estos resultados del experimento son diferentes pero cada uno de ellos tiene la misma probabilidad de ocurrir en el experimento. Su probabilidad de ocurrencia constituye la medición de la incertidumbre del evento, es decir, cuantifica la probabilidad de que ocurra. Como señala Cristófili (2005: 23) “la probabilidad mide la expectativa de que se presente cada uno de los posibles resultados contenidos en el espacio muestral”.

***Definición de probabilidad:* en opinión de Lind et al (2005) es el valor entre 0 y 1 inclusive, que describe la posibilidad (factibilidad o probabilidad) de que ocurra un evento de interés para el investigador dentro de un total de eventos, todos igualmente factibles o con la misma probabilidad de ocurrir.**

Método de Cálculo: Dicho valor se calcula con los métodos *objetivo* (clásico o a priori y/o el de frecuencias relativas o a posteriori) y *subjetivo* (posibilidad de que ocurra un evento en particular, la cual asigna una persona con la información actual disponible a su alcance.).

Reiterando lo anterior pero en forma ampliada, señala el Profesor L. Kazmier (1967) que la teoría de la probabilidad se ha convertido en la base del desarrollo de los métodos que utilizamos en la inferencia estadística, misma que tiene su origen en el método inductivo, el cual indica que a partir del análisis de una porción de eventos o información particular podemos generalizar, es decir, se pasa de lo particular (muestra) a lo general (población o universo); en otras palabras, seleccionamos una muestra (porción) del universo, detectamos **sus características** y decimos que esas mismas características las tiene la población, o sea que la **Inferencia Estadística es** aquella disciplina que basada en el análisis de la muestra por medio de métodos y técnicas científicas, hace posible el conocimiento de las características de la población.

Ahora bien, es importante mencionar que cuando describimos las características de la población, N, a partir de la información de una muestra, n, no estamos seguros de que dicha descripción sea correcta o válida para todos los elementos de la población, por lo que siempre existirá el riesgo de aceptar la descripción cuantitativa de las características de la población a partir de una muestra. Dicho riesgo lo medimos aplicando la teoría de la probabilidad. O sea que en el proceso de información estadística nunca podremos evitar el riesgo o error de aceptar o rechazar a partir de la muestra, características que pueden o no ser ciertas para la población.

Si bien es cierto que no podemos evitar el riesgo, también es cierto que lo podemos controlar y cuantificar por medio de la teoría probabilística.

Idealmente quisiéramos tener a nuestra disposición un *procedimiento de selección de la muestra* que nos garantizara que es representativa de la población para reducir o eliminar el riesgo en la toma de decisiones sobre las características de la población a partir de la información muestral. Desafortunadamente no se ha descubierto tal procedimiento, por lo que nunca estaremos seguros de que una muestra específica sea representativa de una población específica.

Así, en lugar de garantizar que la muestra es representativa, lo mejor que puede hacer el procedimiento de selección es darnos certeza de que no son introducidas fuentes distorsionadoras conocidas durante la selección de la muestra, que en este caso llamamos *muestra probabilística, que,* debe quedar claro, no es necesariamente representativa de la población.

Al respecto, es conveniente decir que uno de los requisitos de una muestra probabilística es que cada elemento de la población estadística tenga una **oportunidad** conocida, generalmente igual, de ser incluido en la muestra.

1 Significado de probabilidad.

Dicha oportunidad se llama probabilidad, la cual podemos **definir como la posibilidad expresada con un número positivo, de que ocurra un evento o resultado de interés para el investigador.** De lo anterior se observa que una expresión probabilística siempre constituye una **ESTIMACIÓN de un valor desconocido que regirá un evento que todavía no ocurre.**

Derivado de lo anterior, cabe mencionar que la teoría de la probabilidad se ha convertido en la base del desarrollo de las técnicas de Inferencia Estadística, las cuales son utilizadas en todos los campos de la investigación básica y aplicada, incluyendo el análisis económico y las decisiones empresariales.

Existen dos procedimientos para el cálculo de la probabilidad:

El 1º se refiere al enfoque objetivo y el 2º se refiere al enfoque subjetivo.

La probabilidad objetiva se calcula por dos métodos:

<u>El clásico ó teórico y el de frecuencias relativas</u>.

El enfoque subjetivo referente a la interpretación de un valor probabilístico, se basa en la confianza o seguridad que una persona tenga sobre la ocurrencia de un evento.

Un ejemplo de éste sería la fuerte creencia, de 0.95, de que se firmará un contrato de la STUNAM y la UNAM. El 31 de octubre del 2002.

Este evento es único, no puede ser repetido muchas veces, sencillamente el 0.95 refleja la confianza que hay sobre la firma del contrato-laboral. De manera general diremos que cuando existe un evento con un sólo resultado posible, el concepto de probabilidad **subjetiva** es aplicable.

Por otra parte, en lo que respecta a la probabilidad **objetiva**, su cálculo por cualquiera de los dos métodos antes mencionados no difiere sustancialmente; su diferencia radica en el **tiempo** en que se calcula determinado valor probabilístico. Esto es, el procedimiento clásico se caracteriza por la determinación **apriorística** de los valores antes de haber observado los eventos; en otras palabras, no es necesario hacer el experimento para observar y registrar su resultado, es decir, la probabilidad se calcula teóricamente.

EJEMPLO:

Cuando se dice que un medio $(\frac{1}{2})$ es la probabilidad de obtener águilas en el lanzamiento de una moneda, esto se dice sin haber lanzado la moneda al aire (el experimento es el lanzamiento de la moneda). Por eso se dice que la probabilidad así calculada **es un valor esperado** con el método clásico o teórico, el cual **supone en** el ejemplo que utilizamos de la moneda, una simetría básica en los posibles resultados de un evento, por ello la moneda o el dado que se utilizarán, no deben estar deformada o en el caso del dado, no debe estar "cargado", para poder calcular la probabilidad a priori.

También debemos decir que el cálculo anterior se basa en el **supuesto** de que los resultados posibles son mutuamente excluyentes e igualmente probables de ocurrir. Al respecto, es conveniente decir que en el mundo de la economía y los negocios los resultados posibles no son igualmente probables y no conocemos de antemano su probabilidad de ocurrencia, situación que limita el uso del método clásico para calcular las probabilidades. La mayor crítica es que el término "igualmente probable" presupone el conocimiento previo de la teoría de la probabilidad, situación que no siempre es cierta, además de que en el mundo real no siempre podemos suponer que los resultados serán "igualmente probables", de ahí que sea interesante, muchas veces, recurrir al método de las frecuencias relativas.

Al respecto, de acuerdo con el método de frecuencias relativas, en que la probabilidad de un evento se basa en un resultado observado o verificado, en otras palabras, las probabilidades se calculan **después** de haber realizado el experimento y una vez que se han registrado los resultados del mismo. Así, la probabilidad de un resultado cualquiera es la frecuencia relativa de ese producto o resultado en un gran número de eventos repetitivos.

Es importante señalar que con este método para calcular la probabilidad, que a medida que **aumenta** el número de observaciones de los eventos y de sus resultados, **aumenta** la exactitud en el cálculo de la probabilidad, inclusive tiende a estabilizarse en cierto valor, por ejemplo, si realizamos el experimento de lanzar al aire, digamos 500 veces una moneda y registramos el número de veces que cae “águila”, la frecuencia relativa, es decir la probabilidad, tiende a estabilizarse alrededor del valor 0.5. Derivado de lo anterior, decimos que la probabilidad así calculada es **un valor estimado**, cuya exactitud será mayor a medida que aumentemos el experimento.

Una vez establecida la diferencia entre uno y otro de los dos métodos del enfoque objetivo, a continuación podemos profundizar señalando lo siguiente:

DEFINICIÓN CLÁSICA DE LA PROBABILIDAD.

Laplace definió la probabilidad como una razón matemática entre un grupo de eventos con características especiales y la totalidad de eventos posibles. Explícitamente diremos: "si un experimento da lugar a **(n)** eventos mutuamente excluyentes, todos igualmente probables y **(r)** se consideran favorables, entonces la probabilidad de un evento favorable es **r/n** ".

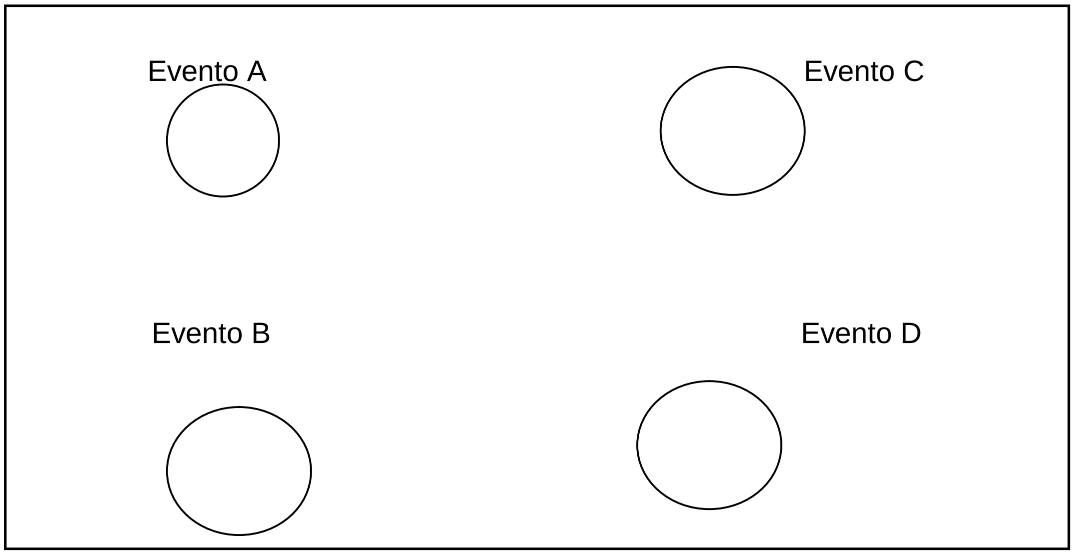
De lo anterior observamos que un valor probabilístico es indicativo de la frecuencia esperada de un resultado posible en particular, dentro del total de resultados posibles que arroje un experimento.

Un evento será una muestra cuyos puntos o elementos son resultados posibles de un experimento. Lo anterior, en el caso de una baraja americana, gráficamente se verá así:

	))))))))))	))))))))))	
• • • • • • • • •	• • • • • • • • • •	• • • • • • • • •	• • • • • • • • • •
• • •	• • •	• • •	• • •
Oooooooooo	ooooooooooo	ooooooooooo	ooooooooooo

Como se observa, un evento puede estar representado por un punto o un agregado de puntos.

Gráficamente:



Serán eventos o resultados verificables A, B, C, D; donde cada uno de ellos está formado por un punto como D, o agregado de puntos como A,B,C.

AXIOMAS DE LA PROBABILIDAD

1º.- A cada punto se le asigna un número positivo, llamado probabilidad.

2º.- Todos los puntos tienen la misma probabilidad de ocurrencia.

3º.- La suma de las probabilidades del espacio muestral es igual a 1.

4º.- La probabilidad de un punto oscila entre 0 y1, es decir $0 \leq p(x) \leq 1$

Conforme a lo anterior podemos establecer que la probabilidad de cada resultado de un experimento es **1/n**.

Al respecto **el espacio muestral** puede definirse como la suma de todos los puntos de una muestra, o de resultados posibles que produce un experimento. En opinión de Yu Lun Chou(21) realmente debería llamarse **"espacio de resultados", porque eso son**.

EJEMPLO.- El experimento "lanzamiento de dos monedas" genera un espacio muestral, conteniendo cuatro puntos o resultados posibles (AA, AS, SA, SS):

donde A = ÁGUILA

S = SOL

El evento "caras iguales", está compuesto de dos puntos (AA y SS). Si queremos saber cuál es la probabilidad de que caigan caras iguales (águilas o soles) en un lanzamiento de dos monedas, con el método clásico, ésta será:

$$P(A,A \text{ ó } S,S) = \frac{\text{Nº de casos favorables}}{\text{Nº de casos posibles}}$$

$$= \frac{2}{4} = \frac{1}{2}$$

Aplicaciones:

La probabilidad fue desarrollada por Pascal

1.- Inferencia estadística:

Muestreo Estadístico, estimación de parámetros y prueba de hipótesis

2.- Econometría:

Modelos Econométricos

3.- Teoría de las Decisiones:

Teorema de Bayes

Para desarrollar la teoría probabilística fue necesario identificar y cuantificar el número de resultados posibles, marco de referencia, y espacio muestral que genera un experimento, puesto que sólo así se puede cuantificar la probabilidad de éxito o fracaso en la obtención de un resultado de interés particular.

Al respecto, la probabilidad se desarrolló en gran parte en los juegos de azar, que constituyen uno de los principales marcos de referencia, la cual posteriormente se utilizó en biología para seleccionar y utilizar muestras probabilísticas que dieran confianza o seguridad a los resultados de sus experimentos. Así, en el caso de la moneda, el marco de referencia son las dos caras de la misma. En el caso de un dado son las seis caras. Cuando son dos dados el espacio muestral está constituido por 36 resultados posibles.

Gráficamente :

D	6	O	o	O	o	o	o	Marco muestral
A	5	O	o	O	o	o	o	constituído por
D	4	O	o	O	o	o	o	36 resultados
O	3	O	o	O	o	o	o	Posibles
#	2	O	o	O	o	o	o	
1	1	O	o	O	o	o	o	
		1	2	3	4	5	6	
		D	A	D	O	#	2	

En el caso de una baraja española, el marco muestral está constituido por 40 cartas o resultados posibles.

En el caso de una baraja americana está constituida por 52 cartas o resultados posibles. Estos resultados se clasifican en 4 grandes grupos: Diamantes, Corazones, Tréboles, Picas, que a su vez se agrupan en dos colores, negro (26 resultados) y rojo (26 resultados).

1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
J	J	J	J
Q	Q	Q	Q
K	K	K	K

Una vez que se conoce el marco de referencia se puede decir qué es posible calcular la probabilidad de ocurrencia de cualquiera de los resultados comprendidos en el marco de referencia. En otras palabras, la probabilidad representa la cuantificación de éxito o fracaso de un resultado posible.

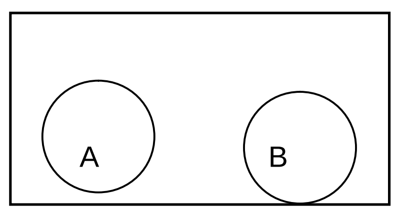
2. RESULTADOS POSIBLES DE UN EXPERIMENTO.

Pueden ser:

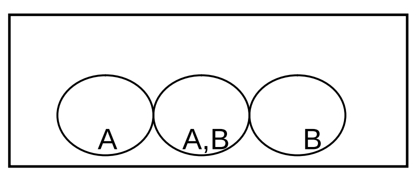
2.1 EVENTOS MUTUAMENTE EXCLUYENTES: A y B lo son cuando en un experimento sólo ocurre uno de ellos. La probabilidad de que ocurra uno o el otro es igual a la suma de sus probabilidades de ocurrencia.

P(A ò B) = P(A) +P(B)

El siguiente diagrama se llama de **diagrama de Venn** y comprende todos los resultados posibles de un experimento, con uno o más resultados identificados específicamente, cuyo conjunto se llama espacio muestral; cualquier resultado se identifica como un punto en ese espacio y el área relativa asignada a ese punto no necesita ser indicativa de su probabilidad.



Cuando hay intersección entre ellos, es decir, que tienen puntos en común, decimos que no son eventos mutuamente excluyentes. Gráficamente se ven así:

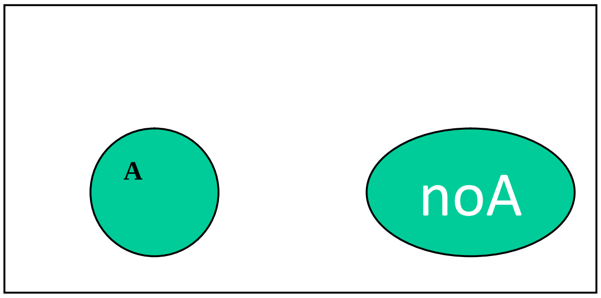


En ese caso el cálculo de su probabilidad es:

P (A ó B)= P(A) + P(B) - P(A,B)

De lo anterior, cuando A y B son mutuamente excluyentes, P(A,B)= 0

En el siguiente diagrama se representa la P(A) y P(no A), ésta última indica la probabilidad de que no ocurra A, tal que P(A) + P(no A) = 1, que ocupan todo el espacio muestral



Los eventos mutuamente **excluyentes** pueden ser más de dos; ejemplo:

Sabemos que la probabilidad de que los estudiantes de posgrado obtengan 10 de calificación es 0.12; P(9)= 0.13 ; P(8)= 0.12; P(7)= 0.18; P(6)= 0.20 y P(5)= 0.25, cuya suma es 1.0, decimos que la suma de todos los resultados mutuamente excluyentes es igual a 1.0, lo cual cumple con uno de los axiomas de la probabilidad. Podemos hacer cálculos como los siguientes:

P(5 o‘6)= 0.25 + 0.20 = 0.45;

P(5 o‘6 ò 7) = 0.25+0.20+0.18 = 0.63;

P(8 ‘0 9) = 0.12+ 0.13 = 0.25,

P (8 ò 9 ò 10) = 0.12+.013+0.12 = 0.37

2.2 EVENTOS INDEPENDIENTES: A y B son independientes cuando ocurren separadamente en el tiempo o en el espacio; cuando la ocurrencia de uno no afecta la del otro. La probabilidad de que ambos ocurran es:

P (A y B) = P (A) * P (B)

Aquí también es conveniente advertir que a diferencia de los resultados posibles que pueden surgir en los juegos de azar, en el mundo de los negocios, los eventos y sus resultados raras veces son independientes, sin embargo, con ese señalamiento, no deja de ser útil para la toma de decisiones.

2.3 EVENTOS DEPENDIENTES Y PROBABILIDAD CONDICIONADA

Cuando A y B no son independientes surge el concepto de probabilidad condicional y para determinar la probabilidad de una **secuencia de eventos** ponemos P(B\ A), significa la probabilidad de que ocurra B dado que A ocurrió previamente.

Ejemplo:

Suponga que un cargamento de diez motores contiene uno defectuoso, D, y nueve no defectuosos, ND. Al inspeccionarlos, obtenga la probabilidad de uno defectuoso, D, y los otros nueve no defectuosos, ND.

Revisión de uno de dos motores

sabemos que para el primero:

P(ND)= 9/10 y que P(D)= 1/10

La revisión de un segundo motor, dado que ya se revisó uno antes puede generar:

1).- P(ND\ND) = 8/9 *9/10 = 72/90= 4/5 =8

2).- P(D\ND) =1/9* 9/10= 9/10=9/90=1/9 =.1

3.- (ND\D)= 9/9*1/10 = 9/90= 1/9= .1

4.- (D\D)= 0

suma = 1.0

Con estas referencias, a continuación recordemos algunos conceptos que necesitaremos para relacionarlos con lo que hemos visto hasta el momento y en su oportunidad dar continuidad al análisis de la relación que tiene la probabilidad con la inferencia estadística.

2.4 FUNCIÓN.- Es una relación de dependencia unívoca de una variable “y” de otra “x”.

Si y = f(x), decimos que los valores de y, variable dependiente, están en función de los valores que tome x, variable independiente.

2.5 VARIABLE.- Es aquella literal (x,y,z, etc.) que toma los valores dados en un espacio muestral dado.

2.6 VARIABLE NUMÉRICA.- (no se considera la probabilidad).**Ahora relacionando lo que conocemos hasta el momento, definamos, calculemos y veamos el alcance de la:**

2.7 VARIABLE ALEATORIA, X,.- Se origina en un experimento aleatorio. Es una función real valorada y definida en un espacio muestral, con su probabilidad de ocurrencia asociada. Así, en el caso de un dado toma los valores: 1,2,3,4,5,6, con su probabilidad asociada de ocurrencia y permite calcular sus **valores esperados: E(Xi)**

E(Xi)= 1(1/6)+2(1/6)+3(1/6)+4(1/6)+5(1/6)+6(1/6) = 21/6= 3.5, que es un valor engañoso, ya que 3.5 no es un valor esperado de Xi puesto que los únicos posibles valores son 1,2,3,4,5,6. Pero es importante conocerlo y saber calcularlo porque **E(Xi) es lo que debe esperarse en un sentido de promedio**, también conocida como la media de Xi = E(Xi)=. «$_{x}$, que también se conoce como media aritmética ponderada en términos de probabilidad, expectativa matemática o media de una variable aleatoria X.

Ejemplo 1:

Sabemos que :

Xi	P(Xi)	Xi- «$_x$	(Xi-«$_x$)^2	P(Xi)* (Xi-«$_x$)^2
1	**1/6**	**-2.5**	**6.25**	**1/6*6.25**
2	**1/6**	**-1.5**	**2.25**	**1/6*2.25**
3	**1/6**	**-0.5**	**0.25**	**1/6*0.25**
4	**1/6**	**+0.5**	**0.25**	**1/6*0.25**
5	**1/6**	**+1.5**	**2.25**	**1/6*2.25**
6	**1/6**	**+2.5**	**6.25**	**1/6*6.25**
suma			**17.50**	**17.50/6**

efectivamente «$_x$ = 1+2+3+4+5+6+/ 6= 21/6= 3.5

Ejemplo 2.

Ahora bien, si el experimento se repite varias veces, no es necesario que el valor esperado sea el un valor posible de la variable aleatoria, como lo muestra el ejemplo anterior de E(Xi)= 3.5. Como concepto, como medida de tendencia central, es un concepto básico que se usa mucho en la economía y los negocios, cuya aplicación en estos campos se ilustra de la manera siguiente:

La probabilidad de que se incendie una casa en la colonia Juárez del Distrito Federal en cualquier día del año 2008, es 0.005. La Compañía de Seguros Monterrey le ofrece al dueño de la casa un seguro contra incendios con una póliza por $ 20,000 por un año; cuyo costo es $ 150.00 .En este caso ¿cuál es la utilidad esperada de Seguros Monterrey?

La utilidad definida por U, es una variable aleatoria que puede tomar los valores de $150.00 si no se incendia la casa y, de $ 19,850.00 si es que se incendia durante el año 2008, periodo que cubre la póliza contratada. Así, la función de probabilidad de U es:

Valor de U: ui	+ $ 150.00	- $ 19,850.00
Probabilidad: P(ui)	0.995	0.005

Su E(Ui) = (150)(0.995) + (-19,850)(0.005)= $ 50.00

La esperanza matemática o utilidad esperada por la póliza vendida siempre debe **ser positiva,** como es el caso, para permitir a Seguros Monterrey el pago de gastos de administración y acumular reservas para pagar los siniestros a los beneficiarios y tenedores de pólizas.

Ejemplo 3.

Lo anterior, desde el punto de vista del comprador, el seguro como cualquier juego de azar que se hace para obtener una utilidad, tiene un valor esperado **negativo.**

Valor de U:ui	+ 19,850.00	- $ 150.00
Probabilidad:P(ui)	0.005	0.995

Su E(Ui)= 19850(0.005) + (-150)(0.995)= 99.25-149.25= $ -50

La cantidad de *menos $50.00 es lo que espera ganar en promedio,* en caso de que se incendie la casa y cobre el seguro por $ 20,000.00.

2.8 VARIABLE ALEATORIA DISCRETA: Es la que toma valores finitos o infinitos numerables, que se pueden contar y que son números enteros.

2.9 VARIABLE ALATORIA CONTINUA: Es la que toma valores infinitos dentro de un intervalo: que son medibles, fraccionables o divisibles.

La varianza de la variable aleatoria se define como $\pm^{2}$= Var(Xi)= E(Xi**-** $«_x$**)^{2}** tanto para variables discretas como continuas. Así, Var(Xi)= 17.50/6=2.91 y la desviación estándar, $\pm = \sqrt{\pm^{2}}$ = 1.70.

Con esos antecedentes, como se indicó, "Laplace" pudo cuantificar la probabilidad al decir que "es una razón" matemática que surge de relacionar un resultado de interés con el total de resultados posibles.

Con literales podemos decir que: **(Cr)** representa uno o varios resultados de interés y **(Cn)** representa el total de resultados posibles, entonces si **(P)** denota probabilidad, por consiguiente la **P(Cr) = r/n**.

Así cuando **r = 1** diremos que la **P(Cr) = 1/n**.

Lo anterior significa que una vez conocido el marco muestral, la probabilidad de cada uno de los resultados posibles tiene la misma probabilidad de ocurrencia, ergo, en el experimento que consiste en lanzar un dado una vez, observamos que la probabilidad de obtener 1, es 1/6, también que la probabilidad de obtener dos es 1/6, y en general la probabilidad de cualesquiera de los 6 resultados posibles es 1/6.

Al respecto es conveniente indicar que para el desarrollo de la teoría probabilística fue válido señalar que todos los resultados posibles en un experimento dado, tienen la misma probabilidad de ocurrencia. No obstante en la vida real esto no es del todo cierto.

3.- ANÁLISIS COMBINATORIO.

En la aplicación de la probabilidad con frecuencia se necesita o es conveniente **contar un gran número de objetos o resultados posibles de un experimento;** en cuyo caso es difícil enumerar o contar el número total de puntos de muestras posibles (subconjuntos) en el espacio muestral, también llamado espacio de resultados. Para resolver esta situación se utilizan las técnicas de **permutación y combinación,** que a su vez, se basan en el **principio**

de multiplicación, el cual establece: " si una operación puede efectuarse en **n1** formas y enseguida, después de realizarse en cualquiera de esas formas, se puede efectuar una segunda operación en **n2** formas, y después de ser ejecutada en cualquiera de estas formas, se puede realizar una tercera operación en **n3** formas, y así sucesivamente hasta k operaciones, entonces las k operaciones pueden ejecutarse en las siguientes formas:

(n1)(n2)(n3) ... (nk-1)(nk) formas

Agreguemos a lo anterior, como referencia adicional, que ya aprendimos a calcular la probabilidad de ocurrencia de los resultados posibles de un experimento, y estuvimos en condiciones de definir y obtener la variable aleatoria, así como su valor esperado o promedio en un espacio muestral determinado.

Importancia del análisis combinatorio.

Ahora vamos a utilizar los conceptos anteriores en el contexto del análisis combinatorio, que a su vez nos permitirá profundizar en la demostración de la relación que tiene la probabilidad con la inferencia estadística, ahora, en el contexto de analizar de cuántas maneras diferentes podemos clasificar o arreglar dichos resultados posibles que, dicho en otras palabras, *podremos saber cuántas muestras podemos obtener y de cuántas maneras distintas podemos constituirlas u ordenarlas con las unidades de muestreo que las componen.*

En general podemos decir que sirve para generar distribuciones probabilísticas y para introducir al lector al muestreo estadístico, en particular al muestreo con reemplazo (permutaciones) y al muestreo sin reemplazo (combinaciones). Su exposición se hace a continuación;

3.1 Permutaciones.

Así, empezaremos diciendo que una permutación es un arreglo de todos o parte de los objetos dentro de un conjunto de objetos en un orden definido. El número total de permutaciones de un conjunto de objetos depende del número de objetos tomados a la vez para cada permutación. El número de objetos tomados a la vez para cada permutación puede ser:

a) Todos los objetos.
b) Parte de los objetos.

Ejemplo del primer caso:

Encontrar el número total de permutaciones del conjunto de letras (a, b, c) tomadas todas a la vez.

Uso del diagrama de árbol: el diagrama de árbol es una gráfica que se usa para mostrar los resultados posibles (permutaciones) cuando éstos se organizan u ordenan por etapas.

Usando el diagrama de árbol, vemos que serían los siguientes:

				PERMUTACIONES
				(Arreglos ordenados posibles)
		b	c	a b c
	a			
		c	b	a c b
O				
r		a	c	b a c
i	b			
g		c	a	b c a
e				
n		a	b	c a b
	c			
		b	a	c b a

Cálculo numérico: $\frac{A}{3} * \frac{B}{2} * \frac{C}{1} = 6$ permutaciones

Hay 6 permutaciones. Nótese que el arreglo A,B,C, es diferente de B,A,C aun cuando cada uno de los 2 arreglos consiste de las mismas letras, luego en este caso decimos que **el orden** en que aparece cada letra es **muy importante**. El número de permutaciones también se puede obtener con la formula:

${}_nP_n$=n!=n(n-1)(n-2)(n-3)...3x2x1= ${}_3P_3$ = 3x2x1 = 6 porque n=3.

Ahora bien cuando n=4 tendremos ${}_4P_4$ = 4! = 4x3x2x1 =24 permutaciones.

EJEMPLO DEL SEGUNDO CASO: SOLAMENTE PARTE DE DOS OBJETOS Si definimos r = el número de objetos, tomados a la vez para cada permutación, entonces la formula es ${}_nP_r$.

${}_nP_r$= el número total de permutaciones de **n** objetos, tomados **r** a la vez. Con n=4 y tomando $_r$ a la vez, calculemos:

a) Tres a la vez; n= 4; r= 3 ${}_nP_r = {}_4P_3$

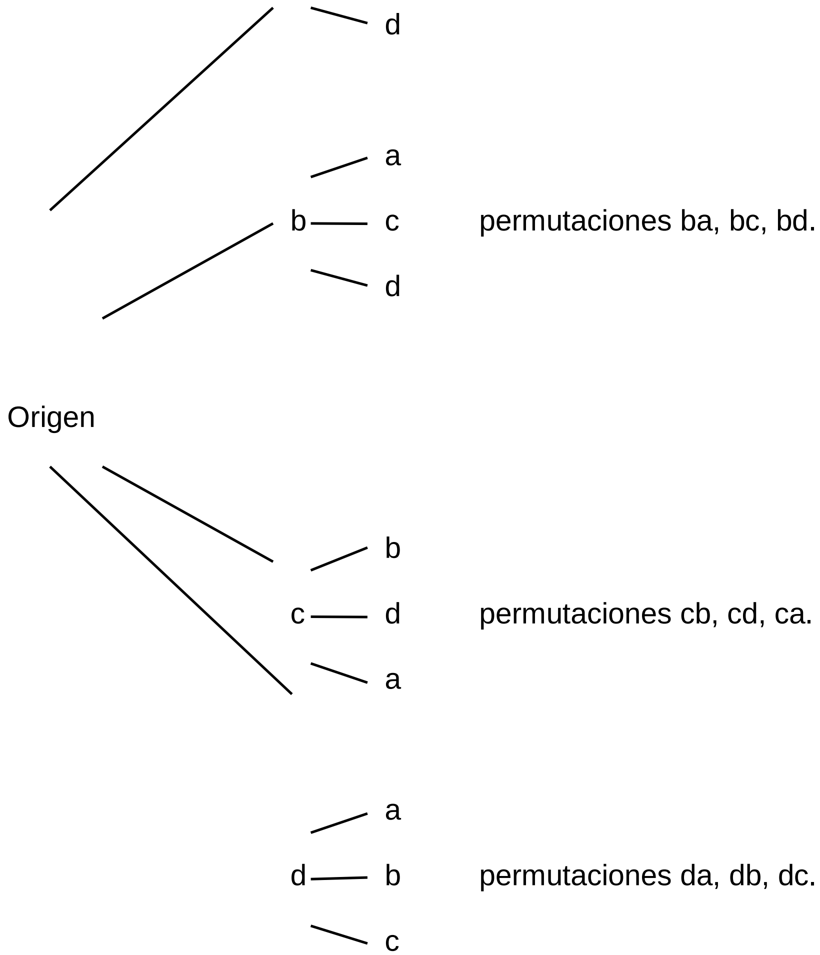
${}_4P_3$ = 4 * 3 * 2 = 24

b) Dos a la vez n= 4 , r= 2 ; ${}_nP_r = {}_4P_2$ = 4*3=12

También se puede obtener con $\frac{n!}{(n-r)!} = \frac{4*3*2*1}{2*1} = \frac{24}{2} = 12$

Lo anterior gráficamente se ve así

$$
\begin{array}{lll}
 & \diagup\ b & \\
a & —\ c & \text{permutaciones ab, ac, ad.}
\end{array}
$$



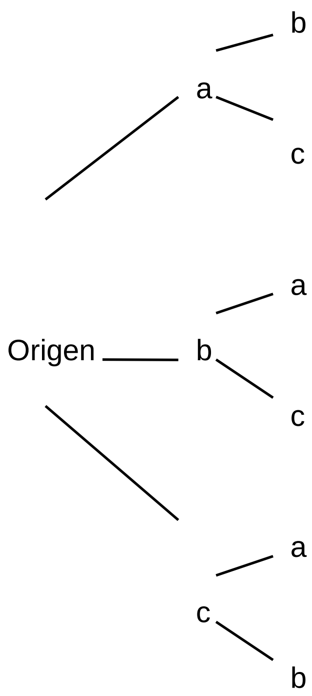
3.2 COMBINACIONES.

Una combinación es un subconjunto o un arreglo de todos o parte de los objetos de un conjunto **sin considerar el orden** de los objetos.

Ejemplo: Encontrar el número total de combinaciones tomando dos a la vez del conjunto (a, b, c).

$\frac{{}_3\boldsymbol{C}_2}{2!} = \frac{3*2}{2*1} = \frac{6}{2} = 3$ combinaciones tomando 2 letras a la vez.

Lo anterior se corrobora usando el diagrama de árbol.



Permutaciones	Combinaciones
ab, ba, ca	ab, bc, ac
ac, bc, cb	

3.3 EJERCICIOS

3.3.1.-DE ANÁLISIS COMBINATORIO AMPLIADO.

Para afianzar el conocimiento, ahora diremos que se utilizan las fórmulas anteriores para obtener numéricamente el número de arreglos diferentes que se pueden obtener cuando ya no es visible el espacio muestral. Supongamos que se tiene **(n)** objetos diferentes y se quiera conocer el número de maneras de ordenar estos objetos. Se puede pensar que hay **(n)** espacios ó lugares donde se puede colocar los **(n)** objetos a fin de dar forma a cada uno de los ordenamientos.

Así habrá **(n)** posibilidades para el primer objeto, n-1 para el segundo, n-2 para el tercero y así sucesivamente hasta llenar el último lugar con el último objeto.

Este desarrollo no es otra cosa que el producto de nPn :

$_nP_r$ = n (n-1)(n-2)...1 = n!; que sería la fórmula para obtener el número total de ordenaciones que también se llaman permutaciones para **(n)** objetos.

En un esfuerzo adicional por consolidar la familiaridad con el manejo de los conceptos que integran el conocimiento del análisis combinatorio, dada la importancia que tiene para la inferencia estadística, decidí complementar mi exposición con la del Profesor S Shao, quien dice:

“Una permutación es un arreglo de todos o parte de los elementos dentro de un conjunto de objetos en un orden definido. El número total de permutaciones de un conjunto de objetos depende del número de los mismos, tomados a la vez para cada permutación puede ser:

a) Todos los objetos ó

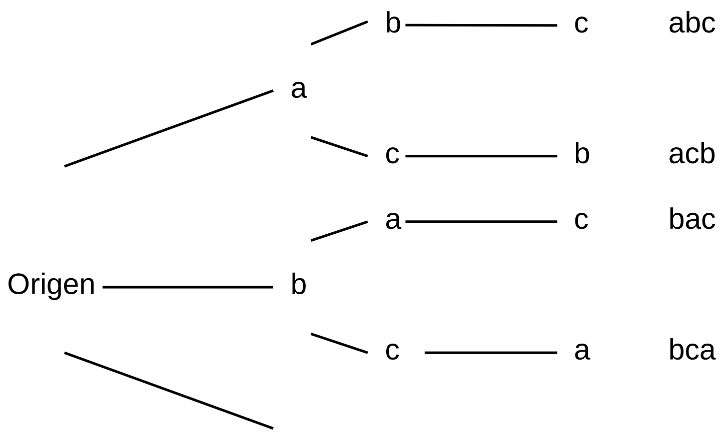
b) Parte de los objetos

Ejemplo del primer caso:

Encontrar el número total de permutaciones del conjunto de letras {a, b, c,} **tomadas todas a la vez**.

PERMUTACIONES

(Arreglos ordenados posibles)



a —— b cab

c

b —— a cba

Hay seis permutaciones. Nótese que el arreglo a, b, c, es diferente de a, c, b, aunque cada uno de los dos arreglos consista de las mismas letras.

El orden de cada arreglo de letras es importante en una permutación. El número de permutaciones se puede obtener con la formula Nº 1.

FORMULA Nº 1.. ${}_nP_r$ = n! = n (n-1)(n-2)(n-3)...3 * 2* 1

Gráficamente será $\frac{\boldsymbol{A}}{3*}\frac{\boldsymbol{B}}{2*}\frac{\boldsymbol{c}}{1}=6$ permutaciones.

También se puede obtener así ${}_3P_3$ = 3!= 3*2*1 = 6 permutaciones.

Otro ejemplo: encontrar el número total de permutaciones del conjunto de dígitos (1, 3, 5, 7,) tomados todos a la vez.

Aquí n = 4 luego 4 P_4 = $\frac{1}{4}\frac{3}{3}\frac{5}{2}\frac{7}{1}=4!=24$ permutaciones, que usando el diagrama de árbol se observa que están ordenadas o integradas de la siguiente forma:

5 ---- 7 -------- 1, 3, 5, 7.

1, 3

7 ---- 5 -------- 1, 3, 7, 5.

3 ---- 7 -------- 1, 5, 3, 7.

1, 5

7 ---- 3 -------- 1, 5, 7, 3.

3 ---- 5 -------- 1, 7, 3, 5.

1, 7

5 ---- 3 -------- 1, 7, 5, 3.

5 ---- 7 -------- 3, 1, 5, 7.

3, 1

7 ---- 5 -------- 3, 1, 7, 5.

1 ---- 7 -------- 3, 5, 1, 7.

3, 5

7 ---- 1 -------- 3, 5, 7, 1.

1 ---- 5 -------- 3, 7, 1, 5.

3, 7

5 ---- 1 -------- 3, 7, 5, 1.

3 ---- 7 -------- 5, 1, 3, 7.

5, 1

Origen

7 ---- 3 -------- 5, 1, 7, 3.

1 ---- 7 -------- 5, 3, 1, 7.

5, 3

7 ---- 1 -------- 5, 3, 7, 1.

1 ---- 3 -------- 5, 7, 1, 3.

5, 7

3 ---- 1 -------- 5, 7, 3, 1.

3 ---- 5 -------- 7, 1, 3, 5.

7, 1

5 ---- 3 -------- 7, 1, 5, 3.

1 ---- 5 -------- 7, 3, 1, 5.

7, 3

5 ---- 1 -------- 7, 3, 5, 1.

1 ---- 3 -------- 7, 5, 1, 3.

7, 5

3 ---- 1 -------- 7, 5, 3, 1.

AHORA PERMUTACIONES DE OBJETOS DIFERENTES TOMADOS PARTE A LA VEZ.

También se puede obtener por medio del diagrama de árbol o con las siguientes fórmulas. El diagrama de árbol es similar a los dos casos anteriores excepto que el número de columnas en este caso es igual al número de objetos tomados para cada permutación. En general sea:

r = El número de objetos, tomados a la vez para cada permutación.

$_n\mathbf{P}_r$ = El número total de permutaciones de n objetos, tomados r a la vez.

Entonces:

Fórmula Nº 2

${}_nP_r$ = n(n-1)(n-2)(n-3)...(n-r+1) para r factores. Nótese que el último factor (n-r+1) es simplificado de [n-r (-1), También cuando r = n, el último factor se vuelve (n-n+1) = 1. Luego cuando r = n, está última fórmula es idéntica a la del número 1.

Ahora bien la fórmula 2 también se puede escribir así:

Fórmula Nº 3: ${}_nP_r = \frac{\boldsymbol{n}!}{(\boldsymbol{n}-\boldsymbol{r})!}$

Esta fórmula es conveniente para cálculos cuando se tiene disponibles tablas de n! y (n-r)!

Ejemplo. Encontrar el número total de permutaciones del conjunto de letras (A, B, C, D) tomados a) tres a la vez y b) dos a la vez.

a) Aquí: n = 4 (Número de letras en el conjunto dado)

r = 3 (Número de letras tomadas a la vez para cada permutación).

${}_nP_r = {}_4P_3 = 4*3*2=24$

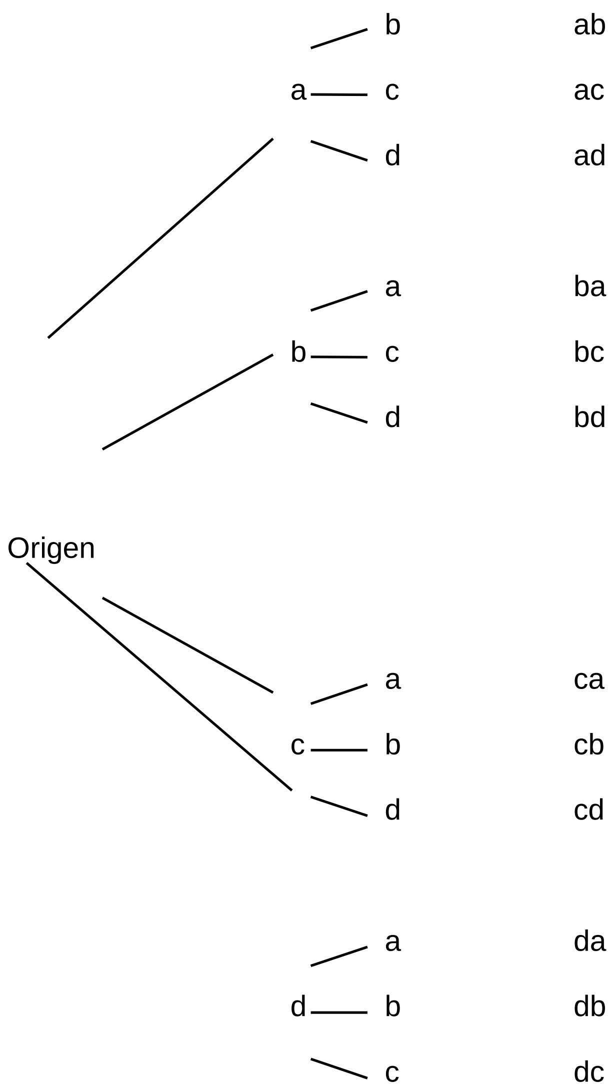
También: ${}_nP_r = \frac{\boldsymbol{n}!}{(\boldsymbol{n}-\boldsymbol{r})!} = \frac{4*3*2*1}{1} = 24\,\boldsymbol{permutaciones}$

c) Aquí, n = 4; r = 2; ${}_nP_r = {}_4P_2 = 4*3 = 12$

También: ${}_nP_r = \frac{\boldsymbol{n}!}{(\boldsymbol{n}-\boldsymbol{r})!} = \frac{4*3*2*1}{2*1} = 12\,\boldsymbol{permutaciones}$

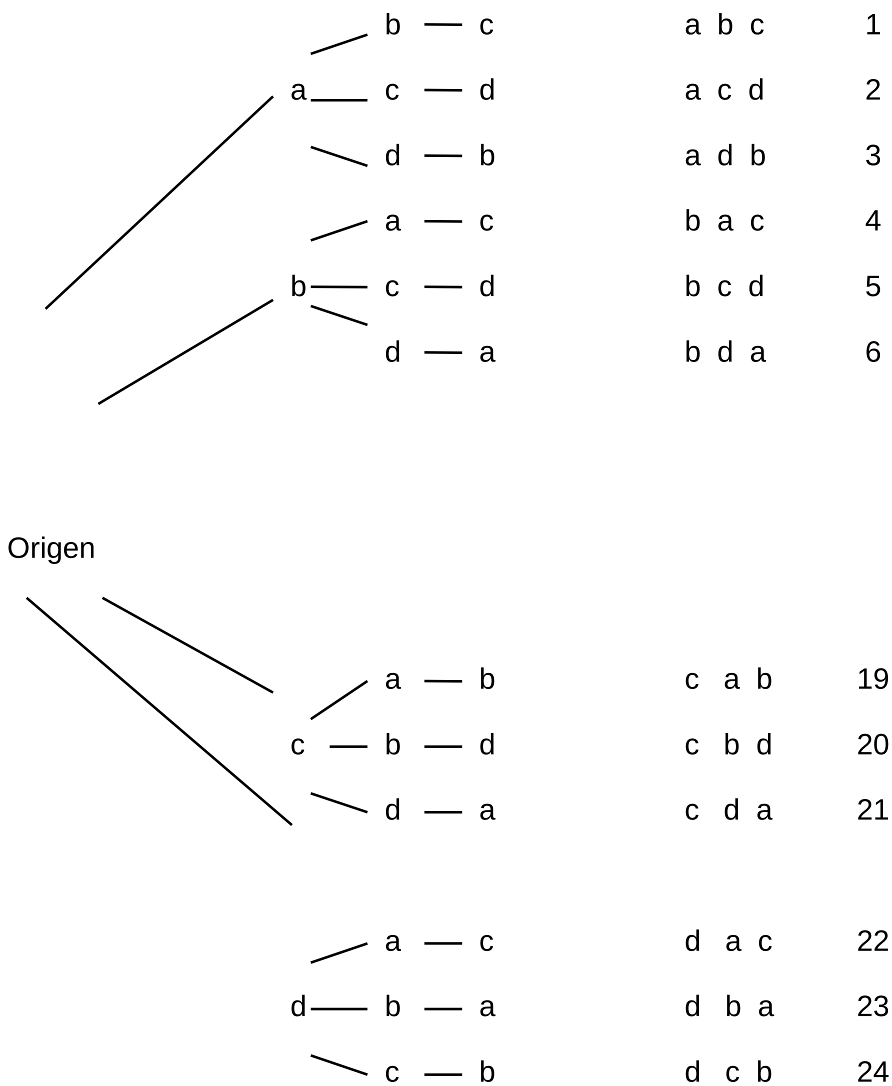
El diagrama de árbol correspondiente se obtiene de la siguiente manera, para las 12 permutaciones:

Resultados posibles



Igualmente, en el caso del inciso a) tendremos:

Cuando ${}_{n}P_{r} = {}_{4}P_{3} = 24$ permutaciones.



Otro ejemplo: Tres oficiales: Presidente, Vicepresidente y Secretario, van a ser elegidos de 20 miembros de un club. ¿De cuántas maneras pueden ser elegidos los tres oficiales?.

Aquí :n = 20; r=3; ${}_nP_r = {}_{20}P_3$ = 20*19*18 = 6840 maneras

ó $${}_nP_r = \frac{\boldsymbol{n}!}{(\boldsymbol{n}-\boldsymbol{r})!} = \frac{20*19*18*17*\ldots 3*2*1}{17*16*15*\ldots 3*2*1} = 6840\,\boldsymbol{\mathit{maneras}}$$

COMBINACIONES.- Es un subconjunto o un arreglo de todos o parte de los objetos de un conjunto sin considerar el orden de los objetos.

El número total de combinaciones posibles de un conjunto de objetos tomados todos a la vez es 1.

Ejemplo:

Los arreglos posibles del conjunto de letras (A, B) son AB y BA. Puesto que el orden del arreglo no es considerado, el arreglo AB es el mismo que BA. Por lo tanto hay solamente una combinación (A y B) posible para el conjunto. Gráficamente:

a ________ b ab

ORIGEN $\begin{bmatrix} dos \\ permutaciones \end{bmatrix}$ y una combinación

b ________ a ba

El número total de combinaciones posibles de un conjunto de objetos diferentes **tomados parte a la vez** puede ser obtenido encontrando primero el número total de permutaciones contando después las permutaciones con los mismos objetos como una combinación.

Ejemplo: Encontrar el número total de combinaciones del conjunto de letras (A, B, C) tomados dos a la vez.

El número total de permutaciones de tres letras, tomadas dos letras a la vez es:

${}_3P_2 = 3*2 = 6$; cada arreglo consiste de dos letras.

Las seis permutaciones consistentes en las mismas letras son consideradas como tres combinaciones. Por lo tanto el número total de combinaciones es:

$$\frac{{}_3P_2}{2!} = \frac{6}{2} = 3$$

USANDO EL DIAGRAMA DE ÁRBOL SE OBTIENE DE LA SIGUIENTE MANERA:

Permutaciones Combinaciones

b ………..ab ……………………………a y b

a

c ……….. ……….ac ……………………a y c

a ………..ba

Origen b

c ……….. ……………….bc …………b y c

a ………………… ca

c

b ……………………………cb

En general, sea:

n = El número total de objetos de un conjunto dado.

r = El número de objetos tomados a la vez para cada combinación.

${}_nC_r$ = El número total de combinaciones de n objetos, tomados r a la vez.

$${}_nC_r = \frac{{}_nP_r}{r!} = \textit{como} \ {}_nP_r = \frac{n!}{(n-r)!} \ \textit{entonces} \ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!}$$

Así el ejemplo anterior también puede ser calculado con estas fórmulas:

${}_nC_r = {}_3C_2 = \frac{3*2}{2*1} = 3 = \frac{n!}{r!(n-r)!} = \frac{3*2*1}{2!(3-2)!} = 3$ combinaciones.

Con ello se puede plantear el siguiente problema: Tres de 20 miembros van a ser seleccionados para formar un comité ¿de cuántas maneras puede ser formado el comité?.

Respuesta: El comité formado por los miembros A, B, C, es el mismo que el comité formado por B, C, A u otro arreglo consistente de los tres miembros, por lo tanto este es un tipo de problema de combinación, puesto que no consideramos el orden del arreglo.

Aquí: n =20; r = 3

$$ {}_{\boldsymbol{n}}\boldsymbol{C}_{\boldsymbol{r}} = {}_{20}\boldsymbol{C}_3 = \frac{{}_{20}\boldsymbol{P}_3}{3!} = \frac{20*19*18}{3*2*1} = 1{,}140 $$

Generalizando entonces tenemos:

COMBINACIONES.

Si hay n objetos diferentes y si deseamos tomar r a la vez, su fórmula será:

$$ {}_nC_r = \frac{{}_nP_r}{r!} = \frac{n!}{r!(n-r)!} $$

De tal forma que si n = 52 y deseamos tomar r = 5 a la vez;

$$ \boldsymbol{P}\begin{bmatrix} 52 \\ 5 \end{bmatrix} = \frac{\boldsymbol{n}!}{\boldsymbol{r}!(\boldsymbol{n}-\boldsymbol{r})!} = \frac{52!}{5!(52-5)!} = 2{,}598.960 $$ maneras o combinaciones.

Si dentro de las 5 cartas, deseamos que 3 sean ases, entonces.

P(3 ases + 2 cartas) = 4(1,128) = 4,512 casos favorables, puesto que

$\begin{bmatrix} 4 \\ 3 \end{bmatrix} = 4$ casos favorables y $\begin{bmatrix} 48 \\ 2 \end{bmatrix} = 1{,}128$ casos favorables, luego

Casos favorables = 4,512 = 0.00174

Casos posibles 2,598,960

3.3.2 EJERCICIOS SOBRE EVENTOS MUTUAMENTE EXCLUYENTES

Hemos dicho que dos o más eventos son mutuamente excluyentes si no puede ocurrir en un cierto experimento más de uno de ellos. La probabilidad de que ocurra uno o el otro dentro de un conjunto de eventos mutuamente excluyentes, es igual a la suma de sus probabilidades de ocurrencia.

Si A = AS ; B = REY

Entonces del ejemplo anterior:

$\boldsymbol{P}(\boldsymbol{B})=\frac{4}{52}=\frac{1}{13};$ también $\boldsymbol{P}(\boldsymbol{A})=\frac{4}{52}=\frac{1}{13};$

Si deseamos conocer la probabilidad de obtener AS o REY, esto es A o B, entonces:

P (A o B) = P (A) + P (B)

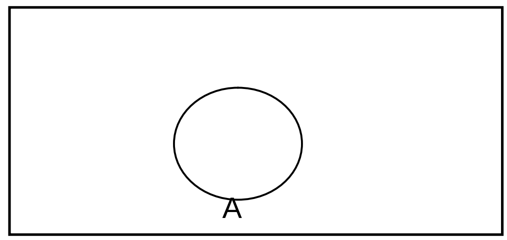
P (A o B) = 1/13 + 1/13 = 2/13

3.3.3 DIAGRAMA DE VENN

Recordemos que un diagrama que comprende todos los resultados posibles de un evento con uno o más resultados específicamente identificados se llama **Diagrama de Venn**.

El conjunto de todos los resultados posibles se llama espacio muestral y cada resultado se identifica como un punto en el espacio.

Utilizando el DIAGRAMA DE VENN: ilustremos la probabilidad de A en un espacio muestral.



Podemos decir que si P (A) es la probabilidad de ocurrencia de A. P (~ A) es la probabilidad de que no ocurra A.

P (A) + P (~ A) =1

En el lanzamiento de un dado la P (AS) es 1/6.

Esto es:

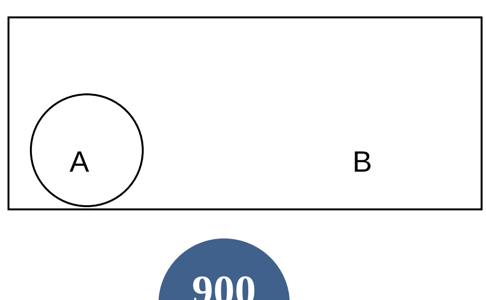
A: P (AS) =1/6

B: P (~ AS) = 5/6

Luego la P (A) + P (B) = 1

Esto es, la suma de las probabilidades de todos los resultados posibles de eventos mutuamente excluyentes es.

1/6+5/6=1



Ejemplos adicionales de eventos mutuamente excluyentes.

1º.- En el lanzamiento de una moneda la ocurrencia de un águila y la de un sol son eventos mutuamente excluyentes.

2º.- El lanzamiento de una moneda dos veces genera eventos mutuamente excluyentes en cada lanzamiento.

3º.- Al sacar una carta de una baraja americana ¿puede salir un as y un rey?

No, luego entonces estos dos resultados posibles son mutuamente excluyentes.

4º.- Al sacar una carta de una baraja americana ¿puede salir un as y una espada?

Si, luego no son eventos mutuamente excluyentes.

El cálculo de los eventos mutuamente excluyentes puede generalizarse para situaciones **en los cuales se manejen 2 ó más eventos mutuamente excluyentes.**

Ejemplo:

No. de hijos por familia	0	1	2	3	4	5 ó más
Proporción	0.10	0.10	0.20	0.25	0.20	0.15

¿Cuál es la probabilidad de que una familia escogida aleatoriamente dentro de un grupo tenga 5 o más hijos?.

Respuesta: 0.15, la proporción representa la probabilidad de acuerdo con el cálculo de la probabilidad por el método de las frecuencias relativas.

¿Cuál es la probabilidad de que una familia tenga tres o más hijos?

A: P (3 hijos) = 0.25

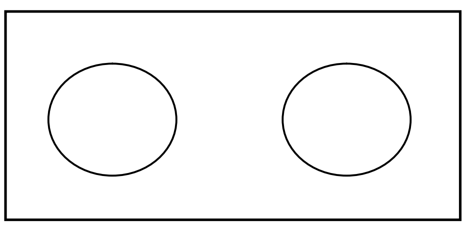
B: P (4 hijos) = 0.20

C: P (5 o más) = 0.15

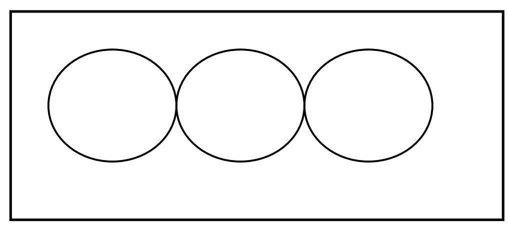
luego: P (A ó B ó C) = 0.25 + 0.20 + 0.15 = 0.60

Si A y B no son mutuamente excluyentes entonces la probabilidad de ocurrencia de A ó B es la probabilidad de que ocurra A más la probabilidad de que ocurra B menos la probabilidad de que ambos ocurran conjuntamente, simbólicamente:

P(A ó B) = P (A) + P (B) - P (AB)



A B



A A,B B

(1)DIAGRAMA DE VENN

ILUSTRANDO 2 EVENTOS

MUTUAMENTE EXCLUYENTES

(2) DIAGRAMA DE VENN

PARA DOS EVENTOS QUE NO

SON MUTUAMENTE EXCLUYENTES.

La sustracción de (A,B) es para corregir el traslape o intersección que se presenta de A y B cuando no son eventos mutuamente excluyentes. Cuando son excluyentes los eventos, A,B = 0, significando que no existe el área (A,B) (en el diagrama 1).

3.3.4 EJERCICIOS SOBRE EVENTOS INDEPENDIENTES.

Ejemplo 1:

Cuando dos o más eventos ocurren en forma secuenciada o separados en el tiempo ó espacio, tales como el lanzamiento de 2 monedas 2 veces, se habla de eventos independientes.

A y B son eventos independientes dentro de un conjunto de eventos si la ocurrencia de uno no afecta la del otro. La probabilidad de que ocurran ambos es P (A y B) = P (A) * P (B).

Ejemplo:

¿Cuál es la probabilidad de obtener dos ases en dos dados en una sola tirada? digamos que A: P (de as en el primer dado = 1/6) y que B: sea la P (de as en el segundo dado) = 1/6)

P (A y B) = 1/6 * 1/6 = 1/36

Son independientes porque un resultado no afecta la ocurrencia del otro.

Ejemplo 2:

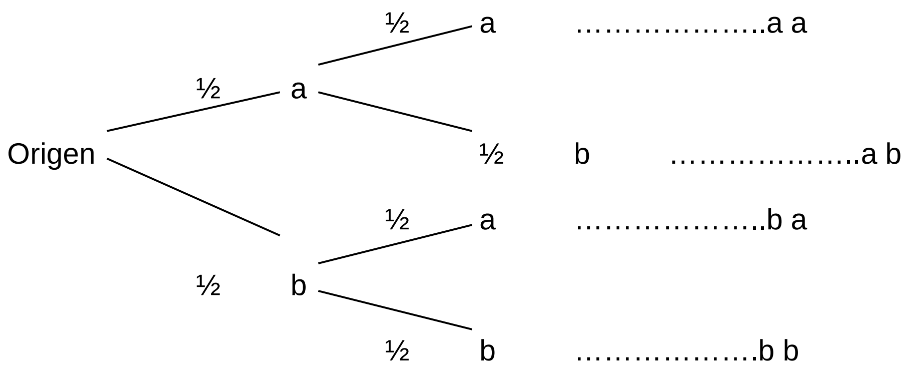
Dos lanzamientos de una moneda son eventos independientes, luego la probabilidad de dos águilas en dos lanzamientos sucesivos de una moneda es 1/4; porque la probabilidad P (A y B) = 1/2 * 1/2 = 1/4; ya que como se recordará P (A y B) = P (A) * P (B).

Por otra parte, es interesante recordar que así como el diagrama de Venn sirve para ilustrar los eventos posibles de un experimento, los diagramas de árbol sirven para ilustrar los resultados posibles de eventos sucesivos o múltiples.

En el caso del lanzamiento de una moneda dos veces el diagrama de árbol será:

a = ÁGUILA b = SOL

Primer lanzamiento Segundo lanzamiento Resultados



¿Cuál es la probabilidad de obtener a y luego b ?

P(A y B) = 1/2 * 1/2 = 1/4

Ejemplo 3: EVENTOS DEPENDIENTES

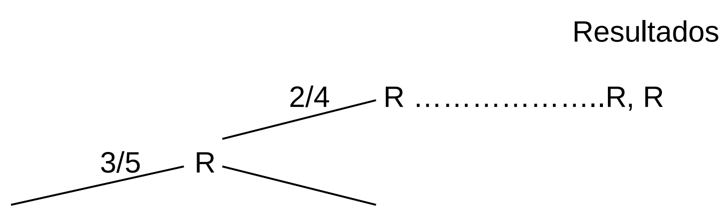
En la vida real la mayoría de los eventos no son independientes, sino que existen interacciones entre ellos. Si son dependientes, el concepto de probabilidad condicionada se usa para determinar la probabilidad de una secuencia particular de eventos, el símbolo P (B\A) significa la probabilidad de B dado que A ocurrió previamente, esto es:

P (A y B) = P (A) * P (B\A)

Ejemplo:

Una caja tiene 3 bolas rojas (R) y 2 negras (N) luego la probabilidad de R = 3/5; P (N) = 2/5 porque son cinco bolas en total.

Si queremos usar el diagrama de árbol, éste será:



Origen $\quad$ 2/4 $\quad$ N ………………..R, N

3/4 $\quad$ R ………………..N, R

2/5 $\quad$ N

1/4 $\quad$ N ………………..N, N

Si en la primera selección obtenemos una bola roja. Obtenga la probabilidad de que en una segunda selección la bola sea negra, sin reemplazo.

P (N\R) = 2/4

P (R y N) = P(R) P(N\R) = $\frac{3}{5} * \frac{2}{4} = \frac{6}{20}$

Por lo tanto

P(RyN)= $\frac{6}{20} = \frac{3}{10}$

Ejemplo 4:

Si la verificación de un evento afecta la probabilidad de ocurrencia de otro, el segundo es un evento dependiente del primero.

Ejemplo:

¿Cuál es la probabilidad de obtener un as en una segunda selección de cartas de una baraja americana?. Ello dependerá de que hayamos escogido un as en la primera selección.

A : P (As en la primera selección es) = $\frac{4}{52} = \frac{1}{13}$

B : P (As en la segunda selección es) = $\frac{3}{51}$

P (A y B) = $\frac{4}{52} * \frac{3}{51} = \frac{12}{2652} = 0.0045$

Ejemplo 5: ***Aplicación de eventos dependientes en economía.***

El cálculo de la probabilidad condicional de un evento dependiente, con un ejemplo económico aplicando el teorema de Bayes o Inferencia Bayesiana, tomado del libro del Prof. J. Kazmier e intitulado "STATISTICAL ANALYSIS FOR BUSINESS AND ECONOMICS de MC Graw Hill, 1967".

El teorema de Thomas Bayes proporciona el procedimiento mediante el cual los valores probabilísticos (a priori) se transforman con base en datos de evidencias actuales en nuevos valores probabilísticos (a posteriori).

Así, suponga que la probabilidad de que nuestro principal competidor decida diversificar su producto es 0.60, y si lo hace hay una probabilidad de 0.80 que construirá una nueva planta.

Así mismo si decide no diversificarse (0.40), hay la probabilidad de 0.40 de que construirá una nueva planta.

Si D = Probabilidad de diversificarse

Si ~D = Probabilidad de no diversificarse

Si B = Probabilidad de construir una nueva planta

Si ~B = Probabilidad de no construir una nueva planta.

Gráficamente podemos ilustrar lo anterior con el diagrama de árbol así:

Diversificarse *Construcción de nueva planta* *Probabilidad condicional*

B P(D) P(B\D) = 0.48

0.80

0.60 —— D

0.20

~B   P(D) P(~ B\D) = 0.12

B   P(~ D) P(B\~D) = 0.16

0.40

0.40 —— ~D

0.60

~B   P(~D) P(~B\~D) = 0.24

Como puede verse B y ~ B dependen de D y son dependientes, su probabilidad esta condicionada a la ocurrencia de D. Así, la probabilidad total de B:

P (B) = P (D) P (B\D) + P (~ D) P (B\~ D)

P= (.6)(.8) + (.4)(.4) = .48 + 0.16 = 0.64

SIMILARMENTE

P (~ B) = P (D) P(~ B\D) + P (~ B) P (~ B\~ D)

P= (.60)(.20) +(.4)(.6) = .12 + .24 = .36

Así P(B ó ~ B) = (0.64) + (0.36) = 1.0

Ahora bien, *si vemos que está construyendo una nueva planta,*

¿Esto indica que ha decidido diversificarse?. No, porque la decisión de construir también pudo haberse tomado con la decisión de no diversificarse.

Luego si deseamos determinar la probabilidad de que nuestro competidor se diversifique dado que está construyendo una nueva planta, usamos el teorema de Bayes, que representa el análisis de la probabilidad condicional cuando se hace una inferencia hacia atrás, es decir se usa en eventos dependientes y de probabilidad condicional, para calcular la probabilidad condicional que permiten hacer inferencias hacia atrás.

De acuerdo con los símbolos usados, para obtener D, se parte de B, llamada probabilidad posterior que sirve para obtener la probabilidad anterior de D, expresada así.

$$P(D \setminus B) = \frac{P(D)\ P(B \setminus D)}{P(B)}$$

P (B) se determina considerando D y ~ D, es decir, cuando se diversifica y cuando no se diversifica. Del diagrama de árbol vemos que:

P(B) = P(D) P(B\D) + P(~ D) P(B\~D) = (0.6)(0.8) + (0.4)(0.4) = 0.64

Luego $\boldsymbol{P}(\boldsymbol{D} \setminus \boldsymbol{B}) = \dfrac{\boldsymbol{P}(\boldsymbol{D})\boldsymbol{P}(\boldsymbol{B} \setminus \boldsymbol{D})}{\boldsymbol{P}(\boldsymbol{D})\boldsymbol{P}(\boldsymbol{B} \setminus \boldsymbol{D}) + \boldsymbol{P}(\sim \boldsymbol{D})\boldsymbol{P}(\boldsymbol{B} \setminus \sim \boldsymbol{D})} = \dfrac{(0.6)(0.8)}{(0.64)} = \dfrac{0.48}{0.64} = 0.75$

Comentarios: Antes de tener la información adicional sobre la construcción de la planta, la probabilidad de diversificarse era de 0.6, que en el lenguaje de la inferencia Bayesiana, se denomina probabilidad apriori. Considerando la información adicional: que nuestro competidor construirá la nueva planta, la probabilidad de que se diversifique ahora es 0.75 y se denomina probabilidad posterior.

La probabilidad posterior puede ser mayor o menor que la apriori. V.gr., si el competidor decidió no construir la nueva planta, la nueva probabilidad posterior de diversificarse sería menor que 0.60.

Demostración:

$$\boldsymbol{P}(\boldsymbol{D} \setminus \sim \boldsymbol{B}) = \frac{\boldsymbol{P}(\boldsymbol{D})\boldsymbol{P}(\boldsymbol{B} \setminus \boldsymbol{D})}{\boldsymbol{P}(\boldsymbol{D})\boldsymbol{P}(\sim \boldsymbol{B} \setminus \sim \boldsymbol{D}) + \boldsymbol{P}(\sim \boldsymbol{D})\boldsymbol{P}(\sim \boldsymbol{B} \setminus \sim \boldsymbol{D})} = \frac{(0.6)(0.2)}{(0.6)(0.2) + (0.4)(0.6)} = \frac{0.12}{0.36} = 0.33$$

Igualmente

$$\boldsymbol{P}(\sim \boldsymbol{D} \setminus \boldsymbol{B}) = \frac{\boldsymbol{P}(\sim \boldsymbol{D})\boldsymbol{P}(\boldsymbol{B} \setminus \sim \boldsymbol{D})}{\boldsymbol{P}(\sim \boldsymbol{D})\boldsymbol{P}(\boldsymbol{B} \setminus \sim \boldsymbol{D}) + \boldsymbol{P}(\boldsymbol{D})\boldsymbol{P}(\boldsymbol{B} \setminus \boldsymbol{D})} = \frac{(0.16)}{(0.64)} = 025$$

0.16+0.48=0.64

$$P(\sim D \backslash \sim B) = \frac{P(\sim D)P(\sim B \backslash \sim D)}{P(D)P(\sim B \backslash D) + P(\sim D)P(\sim B \backslash \sim D)} = \frac{(0.24)}{(0.36)} = 067$$

0.24+0.12=0.36

Ejercicios para reafirmar el conocimiento

1.- ¿Por qué estudiar la probabilidad en economía y en los negocios? ¿Cuál es su importancia ?

2.- ¿La probabilidad permite predecir la ocurrencia de un suceso? ¿Cómo?

3- ¿La probabilidad permite calcular el riesgo o incertidumbre sobre la ocurrencia de un suceso o evento? ¿Cómo?

4.- ¿La probabilidad es el cálculo de que un evento o suceso ocurra en el futuro?

5.- ¿Podemos decir que la probabilidad mide la expectativa de que se presente uno o más de los resultados posibles (suceso o evento) contenidos en el espacio muestral? Explique.

6.- ¿Qué es la probabilidad, cómo la define Laplace y cómo se define en general?

7.- ¿Qué es un experimento?

8.- ¿Qué es un suceso, evento o resultado posible?

9.- ¿Cuántos procedimientos existen para calcular la probabilidad? ¿el subjetivo es uno de ellos?

10.- Dentro del procedimiento objetivo ¿cuántos métodos existen para calcular la probabilidad y cuáles son?

11.- ¿Qué es una variable aleatoria (estocástica), qué es el espacio muestral y qué relación existe entre ellos?

12.- ¿El espacio muestral ilustra gráficamente un proceso estocástico?

13:- ¿Qué es la esperanza matemática y qué relación tiene con la variable aleatoria o estocástica?

14.- ¿Cuáles son los principales axiomas de la probabilidad?

15.- ¿Cuál es la relación de la probabilidad con la inferencia estadística?

16.- ¿Qué es una población y qué es una muestra en sentido estadístico?

|7.-¿Cuál es la diferencia entre una muestra seleccionada probabilísticamente y otra empíricamente?

18.- ¿De qué naturaleza pueden ser los resultados de un experimento?

19.- ¿Cuáles son las fórmulas con que se calcula la probabilidad de ocurrencia de cada uno de ellos?

20.- ¿Para qué sirven las técnicas de permutar y combinar objetos (también llamados eventos, sucesos o resultados posibles de un experimento?

21.- ¿Los resultados de un experimento pueden ilustrarse gráficamente con un diagrama de árbol y el diagrama de Venn? ¿en qué se asemejan y en qué difieren?

22.- ¿Con cuál de los dos diagramas anteriores puede demostrarse fácilmente que los resultados de un experimento pueden ser mutuamente excluyentes e independientes a la vez?

23.- ¿Cuáles son las fórmulas de las permutaciones y de las combinaciones?

24.-¿Por qué las permutaciones y las combinaciones sirven para indicar cómo se selecciona una muestra con y sin reemplazo, respectivamente?

Ejercicio adicional para reafirmar el conocimiento.

PROBABILIDAD DE UNA SUCESO SIMPLE (Salvatore, 1991:26)

A. Método clásico: razón matemática, apriorístico.

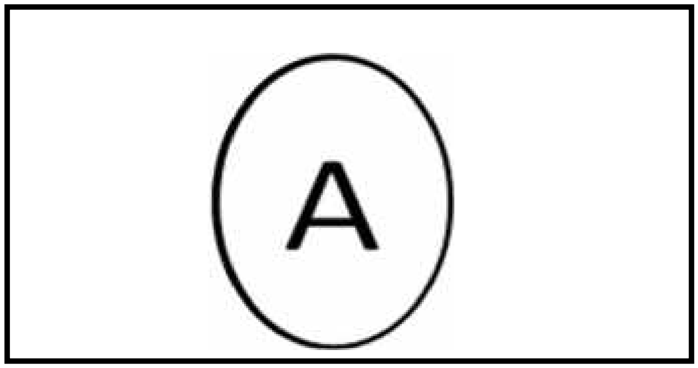
1. Si de un total de N casos posibles en un experimento, todos igualmente factibles, puede ocurrir el evento o suceso A en n_A ,de los casos, tal que la probabilidad de que el evento ocurra está dada por

$P(A) = \frac{n_A}{N}$; donde P (A) = probabilidad de que ocurra A

n_A = número de casos en que A puede ocurrir

N = Número total de casos igualmente posibles.

Dicha probabilidad se visualiza en el diagrama de Venn así:



El círculo representa el evento A y el área total del rectángulo representa todos los casos posibles.

La P (A) varía entre 0 y 1; $0 \leq P(A) \leq 1$.

Cuando P (A) = 0 el evento A no puede ocurrir. Si P (A) = 1 el evento A ocurre con certeza.

Ahora si P (~A) representa la probabilidad de no ocurrencia del evento A.

Luego, P (A) + P(~A) = 1.

2. Si con el lanzamiento de una moneda sin deformaciones se generan dos resultados posibles: (A) ,águila y (S) sol, entonces:

$P(A) = \frac{n_A}{N}$ =1/2

$P(A) = \frac{n_s}{N} = \frac{1}{2}$

Por consiguiente, $P(A) + P(S) = 1$

3. Si el experimento consiste en obtener la probabilidad de los resultados posibles al lanzar una vez un dado, decimos que el dado tiene 6 caras y por ende 6 resultados igualmente posibles, ellos son: 1,2,3,4,5 y 6; tal que

P (1) = $\frac{1}{6}$; P (2) = $\frac{1}{6}$; P (3) = $\frac{1}{6}$; P (4) = $\frac{1}{6}$; P (5) = $\frac{1}{6}$ y P (6) = $\frac{1}{6}$.

Luego P (1) = P (2) = P(3) = P(4) = P(5) = P(6)

así, por ejemplo, la probabilidad de no obtener 3 es P (~3) = 1 – P(3) = 1- $\frac{1}{6}$ = $\frac{5}{6}$, tal que P (3) + P (~3) = $\frac{1}{6}$ + $\frac{5}{6}$ = 1.

4. De los dos últimos ejemplos deducimos que el enfoque clásico para calcular probabilidades parte del supuesto de simetría en la ocurrencia de resultados posibles de un experimento (Kazmier, 1967: 83)

B. Método de frecuencias relativas.

a) Si el experimento consiste en lanzar 100 veces al aire una moneda no deformada y si registramos los resultados posibles: Águila (A) 68 veces y Sol (S) 32 veces, decimos que la frecuencia relativa de águila (A) es 68/100 ó 0.68, que es la probabilidad de ocurrencia de águila (A), que es distinta de la probabilidad a priori o clásica: P (A) = 0.5

b) Si aumentamos el número de lanzamientos de la moneda al aire, observamos que cuando alcanza el infinito en el límite, la frecuencia relativa o probabilidad empírica se acerca a la probabilidad a priori o clásica. Es decir con cualquier método P (A) = 0.5

PROBABILIDAD DE EVENTOS MÚLTIPLES

A. Eventos mutuamente excluyentes (A y B).

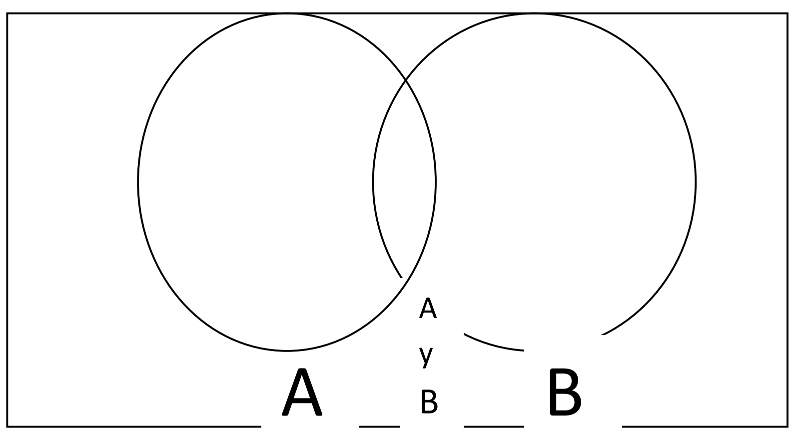
De acuerdo con la definición dada previamente sabemos que

P (A ó B) = P (A) + P (B)

B. Eventos no mutuamente excluyentes.

P (A o B) = P (A) + P (B) – P (A y B)

Usando el diagrama de Venn, lo anterior se expresa así:



C. Eventos independientes.

Así como el diagrama de Venn, ilustra la ocurrencia de un evento, el diagrama de árbol ilustra la ocurrencia de varios eventos o eventos sucesivos.

Con base en la definición previa, decimos que la probabilidad del conjunto P (A y B) = P (A) * P (B).

D. Eventos dependientes: Si A y B lo son, entonces P (A y B) = P (A) P(B/A)

Se dice que la probabilidad de que ocurran A y B es igual a la probabilidad del evento A por la probabilidad del evento B dado que el evento A ya ha ocurrido, dado que

P (B/A) = probabilidad condicional de B dado A.

También P (A y B) = P (B y A).

Ejemplos:

1. Si consideramos el lanzamiento de un dado, el experimento genera los siguientes resultados o eventos mutuamente excluyentes: 1,2,3,4,5 y 6. Así: P (1) = P (2) = P (3) = p (4) = P (5) = P(6) = $\frac{1}{6}$.

Luego P (2 o 5) = P (2) + P (5) = $\frac{1}{6} + \frac{1}{6} = \frac{2}{6} = \frac{1}{3}$.

Generalizando P (1 o 3 o 6) = P (1) + P (3) + P (6) = $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$.

P (1 o 2 o 4 o 6) = P (1) + P (2) + P (4) + P (6) = $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$.

$$P\,(1 \text{ o } 2 \text{ o } 3 \text{ o } 4 \text{ o } 5 \text{ o } 6) = \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} + \tfrac{1}{6} = \tfrac{6}{6} = 1$$

La probabilidad de obtener un as o una espada al sacar una carta de una baraja americana, y sabiendo que P (as) = $\tfrac{4}{52}$ y que P (espada) = $\tfrac{13}{52}$, será P (as o espada) = $\tfrac{4}{52} + \tfrac{13}{52} - \tfrac{1}{52} = \tfrac{16}{52} = \tfrac{4}{13}$

3. ¿Cuáles son las ventajas y desventajas de calcular la probabilidad con: a) el método clásico o apriorístico; b) las frecuencias relativas o probabilidad empírica; c)subjetivamente?

Ventajas:

a) del método clásico: no tenemos que realizar el experimento;

b) de las frecuencias relativas: ellas son las probabilidades de ocurrencia de eventos que sucedieron en el pasado, y

c) subjetivamente se fundamenta en el grado de confianza que una persona tiene de que ocurra un evento.

Desventajas:

a) del método clásico: aún cuando es adecuado en los juegos de azar, en la vida real no es posible, sobre todo en la economía y en los negocios, i.e, es difícil fijar probabilidades a priori de ocurrencia de los eventos de interés y, mucho menos decir que tienen la misma probabilidad de ocurrencia los eventos de interés;

b) de las funciones relativas: se obtienen probabilidades (frecuencias relativas) diferentes para números diferentes de experimentos, i.e., tiene un manejo casuístico que no siempre permite generalizar; además, puede resultar *caro aumentar* el experimento para que estas probabilidades se acerquen a las probabilidades obtenidas con el método clásico;

c) subjetivo o personalista: diversas personas pueden observar la misma realidad con información diferente y por ello, calcular o manejar probabilidades de ocurrencia diferentes para el mismo evento.

4. Dadas las siguientes probabilidades relativas al número adicional de personas en ingeniería que se necesitan en ICA durante los próximos 2 años:

Número de Ingenieros	<100	110-190	200-299	300-399	400-499	≥500
Probabilidad	0.30	0.15	0.30	0.30	0.10	0.05

a) ¿Cuál es la probabilidad de que ICA llegase a necesitar 400 o más ingenieros adecuados en los próximos 2 años?

Si. P (400 – 499) = 0.10

P (≥ 500) = 0.05

Luego P (400-499) ó (≥ 500) = P(400-499) + P(≥ 500) = 0.10 + 0.05 = 0.15

b) ¿Cuál es la probabilidad de que ICA llegue a necesitar al menos 200 pero no más que 399 ingenieros adicionales?

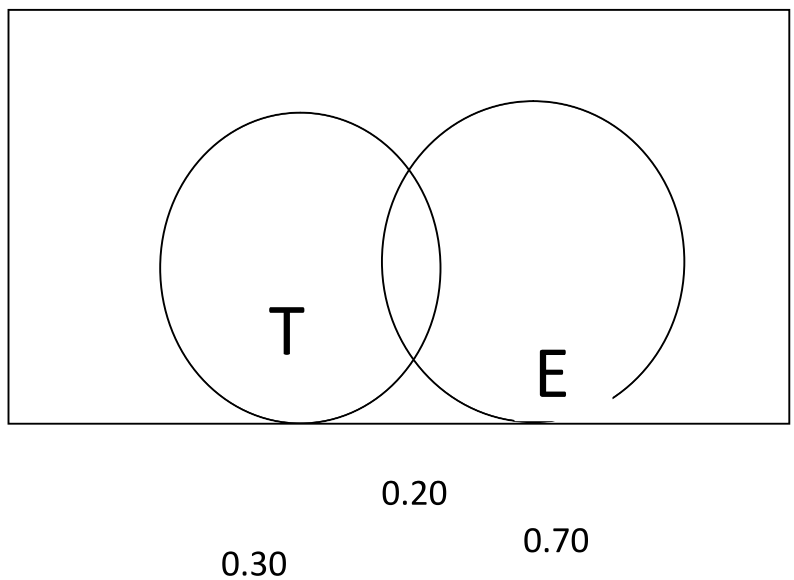
P (200-299) o P (300-399) = P (200-299) + P(300-399) = 0.30 + 0.30 = 0.60

5. Si 0.30 es la probabilidad de que un solicitante de empleo en ICA esté titulado como ingeniero y 0.70 de que haya tenido alguna experiencia como ingeniero y 0.20 de que tenga ambos, ¿de 300 solicitantes qué número de ellos tendrán el título de ingeniero o alguna experiencia de trabajo en ingeniería?

Si T = Titulado y E = Experiencia, decimos que:

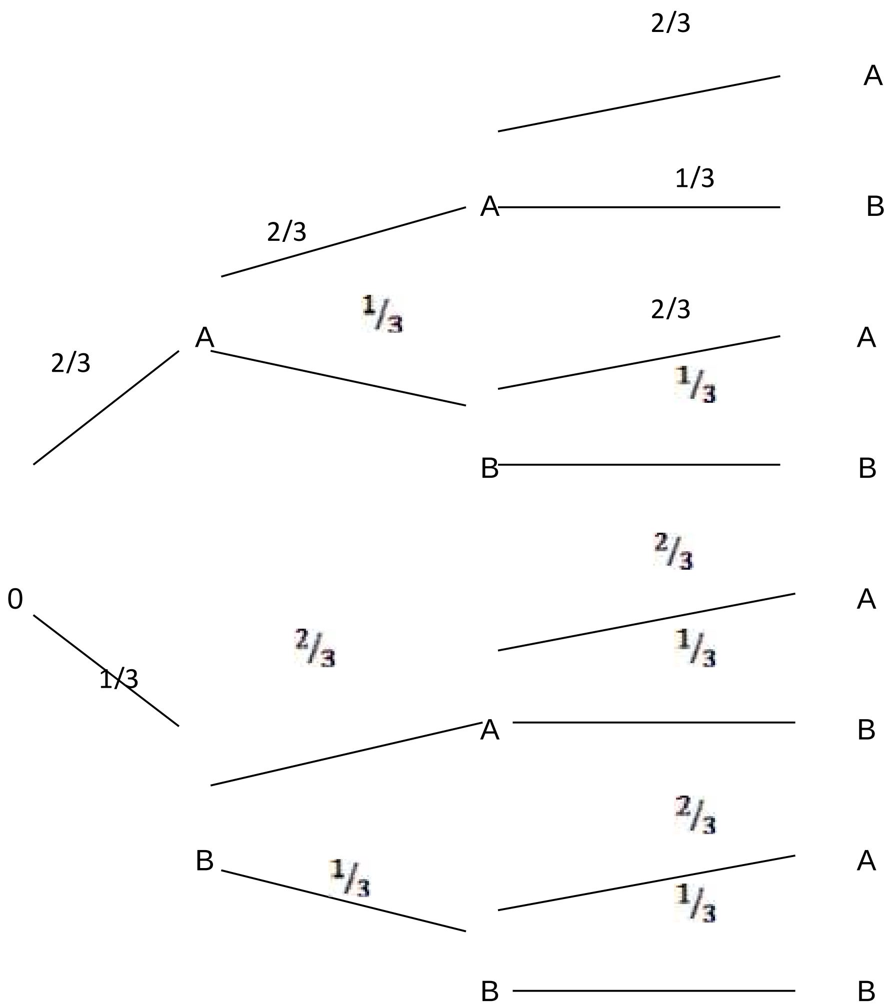
P (T o E) = P (T) + P(E) - P(T, E) = 0.30 + 0.70 - 0.20 = 0.80.

5. Construye un diagrama de Venn para la situación descrita en el problema anterior:



7. Durante el año 2009 dos terceras partes de las acciones preferentes aumentaron sus precios o los mantuvieron estables, en tanto que una tercera parte disminuyó su precio; suponga que analiza la evolución en el mercado de los precios de tres acciones preferentes, seleccionadas al azar: a) Usando A para significar que aumentó o se mantuvo constante el precio de las acciones y B para denotar que su precio disminuyó, construya un diagrama de árbol ilustrándolas probabilidades del aumento o disminución en el precio de las tres acciones seleccionadas al azar (tip, debe de ser un árbol de 3 etapas de izquierda a derecha).

1ra acción 2da. acción 3ª acción.



b) Con los datos del diagrama de árbol anterior, ¿cuál es la probabilidad de que hayan disminuido los precios de las 3 acciones?

P (B, B, B,) = $\frac{1}{3} x \frac{1}{3} x \frac{1}{3} = \frac{1}{27}$

c) ¿Cuál es la probabilidad de que al menos una de las tres acciones haya disminuido en su precio (tip, sólo una rama en el diagrama de árbol no satisface esta condición, y por ende la probabilidad de que los 3 productos secuenciados en este árbol pueda sustraerse de 1.0):

P (al menos una B) = 1- P (A, A, A) = $1-\left(\frac{2}{3}x\frac{2}{3}x\frac{2}{3}\right)=1-\frac{8}{27}$

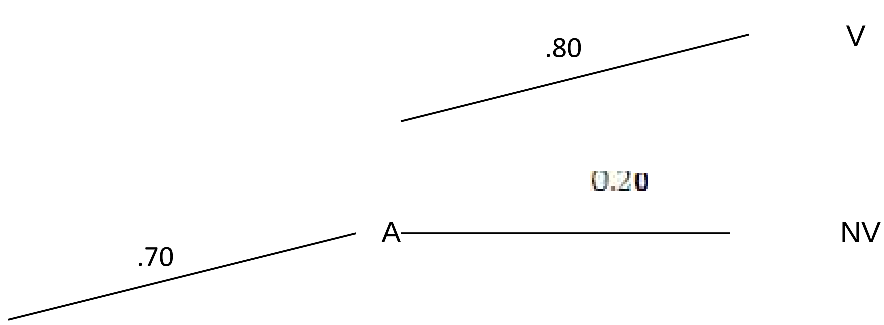
$$\frac{1}{1}-\frac{8}{27}=\frac{27-8}{27}=\frac{19}{27}$$

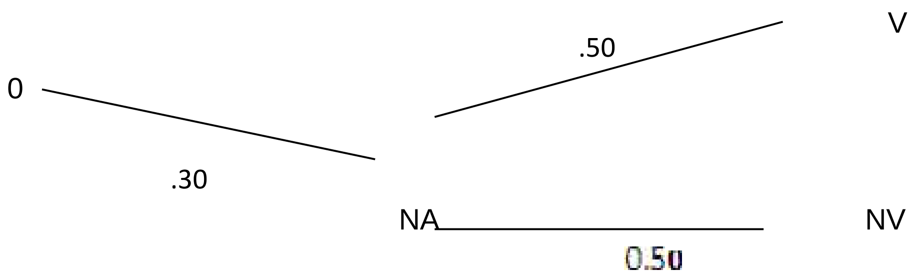
8. La probabilidad de un aumento (A) en la demanda de pan Bimbo para el año próximo se estima que será de 0.70. Si ello sucede 0.80 es la probabilidad de que aumenten las ventas de pan Bimbo. Si no sucede, 0.50 es la probabilidad de que aumenten las ventas de pan Bimbo.

a) Construya el diagrama de árbol ilustrando todos los posibles productos con su probabilidad de ocurrencia asociada usando A y NA para denotar aumento y no aumento en la demanda de pan Bimbo, y V y NV para expresar el incremento y el no incremento en sus ventas, respectivamente.

Empresa: Pan Bimbo

Aumento en su demanda　　　　　　Aumento en sus ventas.





b) ¿Cuál es la probabilidad de que haya un aumento en la demanda y un incremento en las ventas de pan Bimbo?

P (A, V) = P (A) P (V\A) = (0.70)(0.80) = 0.56

c) ¿Cuál es la probabilidad de que ambos no aumenten?

P (NA, NV) = P (NA) P (V\NA) = (0.30) (0.50) = 0.15

9. Los empleados de la empresa Coca Cola Drink van a ser incluidos en el SAR inscribiéndolos en una AFORE. Para ello se toma una muestra de ellos para ser encuestados o entrevistados, mismos que se observa se clasifican de la manera siguiente:

Clasificación:	Evento	No. de empleados
Supervisores	A	120
Mantenimiento	B	50
Producción	C	1460
Administración	D	302
Secretarias	E	68
TOTAL		2000

Vemos que son mutuamente excluyentes (ocurre uno u otro) y colectivamente exhaustivos (por lo menos uno de los eventos puede ocurrir al realizar el experimento). Con estas referencias:

a) ¿Cuál es la probabilidad de que la primera persona seleccionada sea:

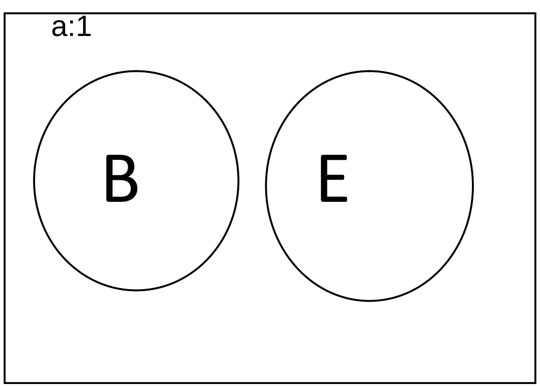
1. Un empleado de mantenimiento o una secretaria

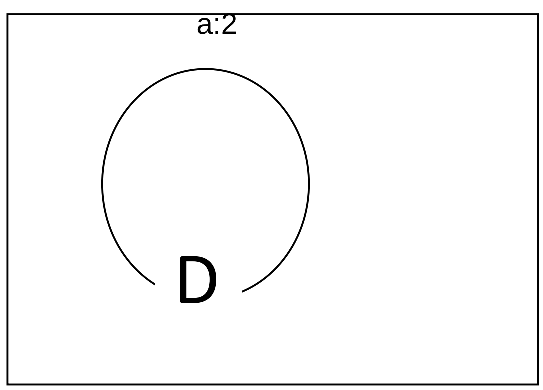
P (mantenimiento o secretaria) = $\frac{50}{2000} + \frac{68}{2000} = \frac{118}{2000} = 0.059$

2. ¿Un empleado (x: administración) que no forma parte de la gerencia?

P(X) = $1 - \frac{302}{2000} = 0.849$

b) Construya un diagrama de Venn que ilustre la respuesta de a).



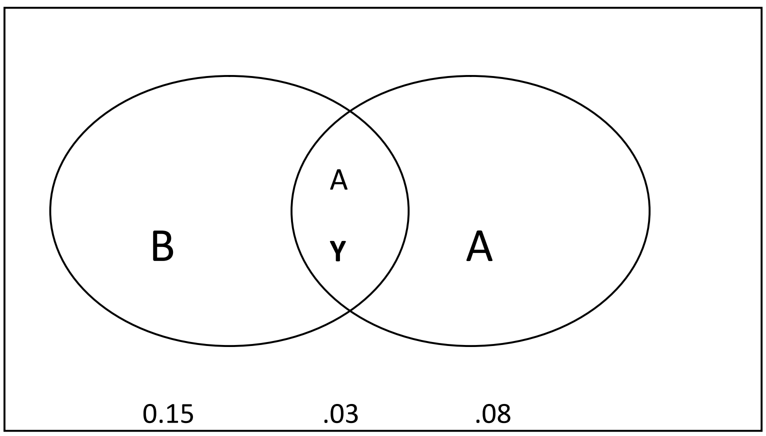


10. En el examen médico anual practicado a los empleados de la empresa “Tequila Jaltotongo” se halló que 8% de ellos necesitan zapatos ortopédicos , 15% requieren de atención dental y 3% necesitan ambos: zapatos ortopédicos y atención dental.

a)¿Cuál es la probabilidad de que un empleado seleccionado al azar necesite zapatos ortopédicos (A) o tratamiento dental (B)?

P (A o B) = P(A) + P (B) – P(A y B) = 0.08 + 0.15 – 0.03 = 0.20

b) Representa lo anterior con el diagrama de Venn.



11. Por su experiencia la fábrica de llantas sabe que la probabilidad de que su marca "Duracero" dure 60,000 km. es 0.80. Si para verificarlo toma una muestra de 4 llantas al azar, ¿cuál es la probabilidad de que las 4 llantas duren 60,000 km.?

(0.80) (0.80) (0.80) (0.80)=0.4096

12. El Consejo de Administración de la empresa "Gatopardo" lo constituyen 8 hombres y 4 mujeres. De entre ellos se debe elegir al azar un comité de búsqueda de 4 miembros para buscar en todo México un nuevo presidente para la empresa.

a) ¿Cuál es la probabilidad de que los 4 miembros del comité de búsqueda sean mujeres?

$$\left(\frac{4}{12}\right)\left(\frac{3}{11}\right)\left(\frac{2}{10}\right)\left(\frac{1}{9}\right)=\frac{24}{11880}=0.002$$

b) ¿Cuál es la probabilidad de que los 4 miembros sean hombres?

$$\left(\frac{8}{12}\right)\left(\frac{7}{11}\right)\left(\frac{6}{10}\right)\left(\frac{5}{9}\right)=\frac{1680}{11880}=0.1414$$

c) ¿La suma de las probabilidades de los eventos descritos en a) y b) es igual a 1?

No, porque hay otros eventos posibles.

13. Sobre el uso de una tabla de contingencia. Sabiendo que es una tabla que se utiliza para clasificar las observaciones de las muestras de acuerdo en dos o más características que se pueden identificar, como la vigente; que es una tabulación realizada que resume al mismo tiempo dos variables de interés y su relación.

Lealtad de los ejecutivos y tiempo de servicio en la empresa "Jugos Naturales del Sureste", denotada con A_1.

Lealtad	Tiempo de servicio en años				
	Menos de 1	De 1 a 5	De 6 a 10	Más de 10	Total
	B_1	B_2	B_3	B_4	
Permanencia en A_1	10	30	5	75	120
No permanencia en A_1: ~A	25	15	10	30	80
Total	35	45	15	105	200

a) ¿Cuál es la probabilidad de seleccionar un empleado con más de 10 años de servicio?

P (B_4)= $\frac{105}{200} = 0.525$

b) ¿Cuál es la probabilidad de seleccionar un empleado que no pertenezca en A debido a que tiene más de 10 años de servicio?

P ($A_2 \backslash B_4$) = $\frac{30}{105} = 0.286$

c) ¿Cuál es la probabilidad de seleccionar un empleado con más de 10 años de servicio o a uno que no permanezca en la empresa A_1?

P (A_2 ó B_4) = $\frac{80}{200} + \frac{105}{200} - \frac{30}{200} = \frac{155}{200} = 0.775$

d) ¿Cuál es la probabilidad de que un empleado permanezca con la empresa A_1?

P (A_1)= $\frac{120}{200} = 0.60$

e) La probabilidad de que un empleado haya trabajado en la empresa menos de un año es P (B)= $\frac{35}{200} = 0.175$

f) ¿Cuál es la probabilidad condicional de que un empleado con más de 10 años de servicio continúe con A_1?

P (B_4\A_1)= $\frac{75}{120} = 0.625$

14. CONCEPTO DE PROBABILIDAD CONJUNTA.

Una probabilidad conjunta es la probabilidad de que dos o más eventos (A y B) ocurran al mismo tiempo. Los eventos A y B no son mutuamente excluyentes. De la tabla de contingencia anterior podemos decir un empleado puede estar dispuesto a permanecer en la empresa y tener menos de un año de experiencia. Esta probabilidad se conoce como probabilidad conjunta y se escribe P (A_1 y B). En la tabla mencionada vemos que hay 10 empleados que se quedarán en la empresa y tienen menos de un año de servicio; por lo que P (A_1 y B) = $\frac{10}{200} = 0.05$ e indica que hay intersección entre ellos por lo que en el cálculo de sus probabilidades como eventos mutuamente excluyentes se debe restar este valor así:

P(A o B) = P(A) + P(B) - P(A y B) = 0.60 + 0.175 – 0.05= 0.725

Interpretación: la probabilidad de que un empleado permanezca en la empresa o haya trabajado por menos de un año es 0.725

15. El método para calcular la probabilidad como una razón matemática, también conocido como clásico o a priori, se aplica cuando hay n resultados igualmente posibles en un experimento.

El método de las frecuencias relativas, llamado también a posteriori o empírico, se aplica cuando se realiza el experimento y se registran los eventos o resultados que éste produce.

16. Teorema de Bayes. Es un método para revisar una probabilidad (a priori) cuando se obtuvo información adicional (a posteriori). Para su exposición antes diremos que un teorema es una suposición científica que demostrarse y que un axioma es una proposición primero o evidencia no susceptible de demostración (Larousse, 2005:973 y 126), y que un axioma es una proposición primera evidencia no susceptible de demostración, que es lógica y se interpreta como un principio enunciado hipotéticamente como base de una teoría deductiva.

Con estas definiciones diremos que este teorema fue desarrollado en el siglo XVIII por el Presbítero Thomas Bayes, quien se preguntó ¿en verdad existe Dios? Y para contestar esta pregunta elaboró una fórmula para llegar a la probabilidad de que Dios existe (Lind et al, 2005: 160), a partir de las evidencias o información a su alcance en la tierra. Luego Laplace detalló la investigación de Bayes y él fue quien la denominó "Teorema de Bayes". Así, para dos eventos mutuamente excluyentes y colectivamente exhaustivos, dicha fórmula es:

$$P(A_1 \backslash B) = \frac{P(A_1)P(B \backslash A_1)}{P(A_1)P(B \backslash A_1) + P(A_2)P(B \backslash A_2)}$$

Para ilustrar su aplicación supongamos que los eventos A_1 y A_2 son mutuamente excluyentes y colectivamente exhaustivos (por lo menos uno de los eventos debe ocurrir al realizar el experimento); además, que A_i se refiere al evento A_1 ó A_2.

Por otra parte suponga que 5% de la población de Sonora tiene una enfermedad que es peculiar en esa entidad federativa. También suponga que A_1 indica el evento "tiene la enfermedad" y que A_2 se refiere al evento "no tiene la enfermedad". Luego entonces si seleccionamos al azar una persona de Sonora, la probabilidad de que esa persona tenga la enfermedad es 0.05 o $P(A_1) = 0.05$. Se conoce como probabilidad a priori (Lind et al, 2005:161). Así se le llama porque se origina antes de obtener cualquier dato empírico (Idem), ya que esta probabilidad inicial se basa en la información actual disponible.

En este sentido apriorístico la probabilidad de que una persona de Sonora no padezca la enfermedad es $P(A_2) = 1 – 0.05 = 0.95$

Ahora bien, sabemos que existe una técnica de diagnóstico para detectar la enfermedad (la cual no es del todo precisa) . Así suponga que B se refiere al evento "las pruebas demuestran que la enfermedad está presente". Suponga también que las evidencias históricas revelan que si una persona tiene la enfermedad, la probabilidad de que la prueba indique su presencia es de

0.90. Así, si utilizamos las definiciones de la probabilidad condicional expuestas previamente, esta afirmación se expresa como: P $(B\backslash A_1) = 0.90$

Ahora suponga que 0.15 es la probabilidad de que una persona que en realidad no tiene la enfermedad la prueba indicará la presencia de ésta. P $(B\backslash A_2) = 0.15$

Con estas referencias ahora suponga que seleccionamos al azar a una persona de Sonora, realizamos la prueba y ésta indica que la enfermedad está presente. Por consiguiente, ¿qué probabilidad hay de que la persona realmente padezca la enfermedad? En forma simbólica, queremos saber P (A_1 dado B) que se lee: P (tiene la enfermedad dados los resultados de la prueba son positivos). La probabilidad P (A_1 dado B) se conoce como probabilidad a posteriori (probabilidad revisada con base en datos adicionales).

Aplicando la fórmula del Teorema de Bayes se determina la probabilidad a posteriori:

$$P(A_1\backslash B) = \frac{P(A_1)P(B\backslash A_1)}{P(A_1)P(B\backslash A_1)+P(A_2)P(B\backslash A_2)}$$

$$= \frac{(0.05)(0.90)}{(0.05)(0.90)+(0.95)(0.15)} = \frac{0.0450}{0.1875} = 0.24$$

Interpretación: 0.24 es la probabilidad de que una persona tenga la enfermedad debido a que la prueba dio positivo. Entonces, si se selecciona una persona de Sonora al azar, la probabilidad de que padezca la enfermedad es 0.05. Si esa persona se somete a la prueba y el resultado es positivo, decimos de que la probabilidad de que realmente esté enferma aumenta de 0.05 a 0.29, es decir, aumenta casi cinco veces. Con estos cambios los cálculos se resumen en:

Evento A_i	Probabilidad anterior	Probabilidad condicional	Probabilidad conjunta	Probabilidad posterior
	P (A)	P (B\A)	P (A_i y B)	P (A_i\B)
Enfermedad, A_1	0.05	0.90	0.0450	0.0450/0.1875 = 0.24
Sin enfermedad, A_2	0.95	0.15	0.1425	0.1425/0.1875 = 0.76
			P(B) = 0.1875	1.00

3.6.- PRÁCTICA

NOMBRE____ GRUPO_____

PROBLEMA 1.- Al mercado concurren tres empresas con los productos A, B, C,. El número de unidades de A es de 20, el de B es de 35 y el de C es de 45. Una unidad será elegida al azar entre todas ellas.

1. ¿Cuál es el conjunto de eventos elementales o espacio muestral?
2. ¿Cuál es la probabilidad asociada a cada evento elemental?
3. ¿Cuál es la probabilidad de elegir una unidad del producto A?
4. ¿Cuál es la probabilidad de elegir una unidad del producto B?
5. ¿Cuál es la probabilidad de elegir una unidad del producto C?
6. ¿Cuál es la probabilidad de elegir una unidad sea del producto A o B?
7. ¿Cuál es la probabilidad de elegir una unidad sea del producto B o C?
8. ¿Cuál es la probabilidad de elegir una unidad sea del producto A o C?

PROBLEMA II.- En una localidad de 10,000 compradores las opiniones respecto a dos productos X y Z se manifiestan de la siguiente manera:

1,000 son favorables a ambos.

2,000 a favor de X y en contra de Z.

1,000 en contra de ambos.

4,000 a favor de X y no tienen opinión sobre Z.

1,000 en contra de Z y no tiene opinión respecto a X.

1,000 no tienen opinión respecto a ambos.

Si se elige al azar un comprador, ¿Cuál es la probabilidad de que?:

1. Opinen a favor de X.
2. Opinen en contra de X.
3. No tiene opinión respecto a X.

PROBLEMA III.- Dentro de una rama industrial se encuentran 15 empresas divididas en tres grupos: grupo México con 6, grupo Puebla con 4 y grupo Querétaro con cinco. Si denotamos por M, P, y Q como los eventos de exportar una misma mercancía, determinar las probabilidades siguientes:

1. Sea una empresa del grupo México la que exporte.
2. Sea una del grupo Puebla la que exporte.
3. Sea una del grupo Querétaro la que exporte.
4. Que no sea del grupo México.
5. Que sea del grupo México o Puebla.

PROBLEMA IV.- En una Facultad de Ciudad Universitaria asisten 2,500 estudiantes con las siguientes características:

1,000 son del sexo femenino.

1,200 pesan 58 kilos o más.

De las mujeres 700 miden sobre 1.58.

De los hombres 1,300 miden sobre 1.65.

De los 2,500 uno se elige al azar:

1. Determinar el conjunto de eventos elementales o marco muestral.
2. ¿Cuál es la probabilidad de elegir un estudiante varón?
3. ¿Cuál es la probabilidad de elegir un estudiante que pese menos de 58 kilos?
4. ¿Cuál es la probabilidad de que habiendo elegido a un estudiante varón, este mida sobre 1.65 metros?

3.7.- EXAMEN: INTRODUCCIÓN A LA PROBABILIDAD

Nombre del alumno:____

1.- ¿Qué es la probabilidad? ____

____;

2.- ¿Cuántos y cuáles son los enfoques calcular la probabilidad?__________

____

3.- Según Laplace, ¿cómo se define la probabilidad?____________________

____

4.- ¿Cuándo se aplica el enfoque subjetivo para calcular la probabilidad?_____

____

5.- ¿En que difiere el método de las frecuencias relativas del método teórico?__

____
____

6.- ¿Para qué sirven las técnicas del análisis combinatorio? _______________

____

7.- ¿Un evento sólo puede estar constituido por un punto? SI_____; NO______

8.- ¿Por qué las permutaciones sirven de referencia para el muestreo con reemplazo?____
____

9.- ¿Por qué las combinaciones sirven de referencia para el muestreo sin reemplazo?____
____

10.- Las permutaciones, ¿son más o menos que las combinaciones, porqué?____

____

Observaciones: Cada una de las respuestas cuenta como medio punto.

PROBLEMA 1:

Referencias: La Secretaría de Economía puede investigar en las empresas zapateras: **a,b** y **c** si éstas cumplen con las normas de calidad que exige el TLC.

Si decide investigar en dos de ellas lo anterior:

a).- ¿Cuántas y cuáles permutaciones (muestras) tiene a su disposición? ____;

b).- ¿Cuántas y cuáles combinaciones (muestras) tiene a su disposición?____;

c).- En términos de representatividad de la muestra, ¿escogería muestras con o sin reemplazo, por qué? ____

____.

PROBLEMA 2:

Si una muestra de empleados de TELMEX participa en una encuesta sobre un nuevo plan de pensiones y si estos empleados se clasifican como se indica en el siguiente cuadro:

Clasificación	Evento	Número de empleados
Supervisores	A	120
Mantenimiento	B	50

Producción	C	1460
Administración	D	302
Secretarias	E	68

a)- ¿Cuál es la probabilidad de que una persona seleccionada al azar sea un empleado de Mantenimiento o una Secretaria?___________________________

b).- ¿Cuál es la probabilidad de que esa persona seleccionada al azar no sea de Administración? ____

c).- ¿Los eventos de la pregunta a) son independientes, mutuamente excluyentes o ambos?____

PROBLEMA 3:

Si usted sabe que cada año a los empleados de la Facultad de Economía les es practicado un examen físico para conocer su estado de salud y que el año pasado se detectó que 8% de ellos necesitaban zapatos ortopédicos, que 15% requieren de un tratamiento dental y que 3% de ellos requieren tanto de zapatos ortopédicos como de servicio dental, así:

a).- ¿Cuál es la probabilidad de que un empleado seleccionado al azar necesite zapatos ortopédicos o tratamiento dental?

b).- Represente esta situación con un diagrama de Venn.

Observaciones: el problema uno cuenta 2 puntos; el dos, 1.5 puntos y el tres, 1.5 puntos.

Examen Número dos

1.- ¿Qué es probabilidad?

Referencias para lo siguiente: El Centro de Diversiones Six Flags hace poco incorporó un nuevo *juego de video* en sus instalaciones. Para probar su aceptación en el mercado se ha invitado a 80 jugadores veteranos para que lo usen y den su opinión sobre su mercado potencial.

2.- ¿Cuál es el experimento?

3.- ¿Cuál es un resultado posible?

4.- Si 65 de los jugadores probaron y dijeron que les gustaba el nuevo juego, ¿65 es una probabilidad?

5.- La probabilidad de que el nuevo juego sea un éxito se calcula en –1. Haga los comentarios pertinentes.

6.- Especifique un evento posible.

Hablando de los métodos para calcular la probabilidad,

7.- En una baraja americana que consta de 52 cartas, ¿Cuál es la probabilidad de que al sacar una carta, ésta sea una reina?

8.- El Instituto de Protección a la Niñez comenta que el estado civil de los padres de familia de los 539 niños inscritos la semana pasada es el siguiente: 333 son parejas casadas; 182 divorciados y 24 son padres viudos, ¿Cuál es la probabilidad de que un niño en particular elegido al azar tenga un padre divorciado?

9.- Con base en sus fuentes de información disponibles ¿Cuál es la probabilidad de que el barril de petróleo tenga un precio mayor a 100 dólares en los próximos seis meses?

Hablando de las reglas para calcular la probabilidad de ocurrencia de dos o más eventos, si una muestra de empleados de SORIANA participa en una encuesta sobre un nuevo plan de pensiones y sí estos empleados se clasifican como se indica en el siguiente cuadro:

Clasificación	Evento	Número de empleados
Supervisores	A	20
Mantenimiento	B	40
Ventas	C	1460
Administración	D	152
Secretarias	E	28

10.- ¿Cuál es la probabilidad de que una persona seleccionada al azar sea un empleado de Mantenimiento o una Secretaria?

11.- ¿Cuál es la probabilidad de que esa persona seleccionada al azar no sea de Administración?

12.- Ilustre con un diagrama de Venn las dos respuestas que dio en 10 y 11.

13.- ¿Los eventos de la pregunta 10 son complementarios, mutuamente excluyentes o ambos?

Si usted sabe que cada año a los empleados de la Facultad de Ingeniería les es aplicado un examen físico para conocer su estado de salud y, que el año pasado, se detectó que 18% de ellos necesitaban lentes pata leer, que 25% requieren de un tratamiento físico para bajar de peso y que 3% de ellos requieren tanto de lentes como de ejercicio, así:

14.- ¿Cuál es la probabilidad de que un empleado seleccionado al azar necesite lentes para leer o hacer ejercicio para bajar de peso?

15.- Represente esta situación con un diagrama de Venn.

16.- La empresa “Llantas para siempre” dice que la probabilidad de que su llanta QR-902 dure en operación 60,000 kilómetros en uso, es 0.80 antes de que se haga lisa. Si usted compra cuatro llantas QR-902, ¿Cuál es la probabilidad de que las 4 llantas duren 60,000 kilómetros en uso?

17.- Por otra parte, sabemos que la Junta de Gobierno de la UNAM consta de 8 hombres y 4 mujeres. Si seleccionamos al azar un Comité de 4 personas para evaluar la educación, ¿Cuál es la probabilidad de que los cuatro miembros del Comité sean mujeres?

18.- ¿Cuál es la probabilidad de que los 4 miembros del Comité sean hombres?

19.- ¿La suma de las probabilidades de los eventos descritos en 17 y 18 es igual a 1?

20.- Si tengo las letras a, b, c, y d y deseo clasificarlas con *cierto orden,* Si tomo dos a la vez, ¿De cuántas maneras puedo clasificarlas?

21.- Ilustre con el diagrama de árbol cada uno de los resultados anteriores.

22.- Si una población (N) consta de los elementos 1,2, y 3 determine su media aritmética (μ).

23.- Si de 22, tomo muestras (n) de tamaño 2 en un muestreo sin reemplazo, ¿Cuántas y cuáles son las muestras a mi disposición para hacer una investigación?

24.- De 23, obtenga la Esperanza matemática de todas las medias muestrales y demuestre que es igual a μ.

25.- De 22, si tomo muestras de tamaño 2 en un muestreo con reemplazo, ¿Cuántas y cuáles sin las muestras que se obtienen?

26.- De 25, ¿Son iguales su esperanza matemática y la media del universo?

27.- De 23, ¿Cuáles son las medias aritméticas muestrales cuyo valor no es igual al valor de la media aritmética de la población? ¿En esos casos cuál es el error de muestreo?

28.- De 23, si las medias aritméticas de las muestras constituyen una distribución de medias, ¿Qué valores puede tomar la variable aleatoria?

II.- **DISTRIBUCIONES PROBABILÍSTICAS**

Este tipo de distribuciones son muy importantes porque una vez conocidas sus características estadísticas y el alcance de cada una de ellas, ampliamos nuestra capacidad de análisis, ya que a partir del conocimiento de sus supuestos teóricos, de su conformación, de su distribución y de la destreza que desarrollemos para saber aplicarlas o adaptarlas a fenómenos económicos específicos, podemos hacer estimaciones de riesgo o incertidumbre, de parámetros, de verificación de hipótesis de trabajo, calcular y utilizar tamaños de muestras para inferir las características de la población de donde las sacamos, etc. Todo ello a partir de muestras sin tener que estudiar toda la población, como sería a través de un censo.

Para saber cómo se generan, empezaremos haciendo el símil con una distribución o arreglo de datos en lo que hemos dado en llamar una distribución de frecuencias, que es una lista de todos los resultados posibles con la asociación de una frecuencia observada por cada resultado.

Similarmente, una distribución probabilística también es una lista de todos los resultados posibles, pero en lugar de la frecuencia observada, se indica la probabilidad asociada con cada uno de los resultados.

Así, si tres monedas se lanzan al aire y se registran los resultados, el número posible de águilas en un lanzamiento puede ser: 0, 1, 2, 3.

Aún cuando hay cuatro resultados posibles sólo uno ocurre en el lanzamiento de tres monedas.

Suponiendo que realizamos o repetimos el experimento de lanzar diez veces las tres monedas y se registra el número de veces que cae 0, 1, 2, 3. la tabla que resulta es una distribución de frecuencias.

No de águilas	Frecuencia Observada
0	2
1	4
2	4
3	0

Si el experimento se repite, una y otra vez, en cada ocasión se obtienen resultados diferentes. Para evitar lo anterior y no conducirnos casuísticamente, es decir, estar tabulando las frecuencias de ocurrencia de cada resultado posible, en forma aislada para luego llegar a conclusiones circunstanciales o coyunturales en el estudio de un fenómeno económico, *es preferible tratar de generalizar* aplicando procedimientos estándar de aceptación general en el análisis de los mismos, cuyos resultados sean creíbles puesto que se maneja una metodología aceptada por la mayoría. Para ello qué mejor referencia que el enfoque clásico o teórico, con el que podemos determinar e indicar la probabilidad de cada producto: 0.1.2.3, ya que en su lugar determinamos o indicamos la probabilidad de cada producto, situación que nos evita que cambie la distribución, es decir, en el caso del experimento de lanzar tres monedas al aire y registrar sus resultados, teóricamente éstos siempre serán: 1/8 para cero águilas o tres soles; 3/8 para un águila y dos soles; 3/8 para dos águilas y un sol y 1/8 para tres águilas.

Reiterando, mientras que una distribución de frecuencias lista todos los resultados posibles con su frecuencia asociada indicando el número de veces que ocurre cada resultado, la distribución probabilística también lista todos los resultados posibles con su probabilidad asociada de ocurrencia, así: partiendo de la definición clásica la cual establece que p = 1/2 = q; donde p = Probabilidad que caiga "águila" y q = Probabilidad de que no sea "águila"; si lanzamos tres monedas a la vez y registramos el número de águilas, generamos una **distribución probabilística** con ocho resultados posibles, que agrupados dan:

No de águilas	Probabilidad	
0	1	8
1	3	8
2	3	8
3	1	8

Uno de los primeros beneficios de estos cálculos es que dada una distribución probabilística, se puede desarrollar una distribución de frecuencias esperadas multiplicando el valor de cada una de las probabilidades por el número total de veces que se repita el experimento. Si lo hacemos 24 veces:

No de águilas	Frecuencia esperada en el lanzamiento de 3 monedas 24 veces					
0	24	*	1	8	=	3
1	24	*	3	8	=	9
2	24	*	3	8	=	9
3	24	*	1	8	=	3

Raras veces la distribución de frecuencias observadas coinciden con la de las esperadas, *que se convierten en la mejor estimación de las primeras si el experimento se realiza muchas veces*. Luego una distribución de frecuencias esperadas es una distribución probabilística.

SU NATURALEZA Y FORMA DE GENERARLAS.

Pueden ser discretas y continuas. Dentro de las primeras destacan por su uso en la economía la distribución binomial, la hipergeométrica y la de Poisson. Dentro de las continuas, la principal y de mayor uso es la distribución normal. Aún cuando existen diferentes maneras de generar una distribución de frecuencias esperadas discreta, dos son usadas extensamente en la inferencia estadística partiendo de la definición clásica de probabilidad: el diagrama de árbol y la expansión del binomio, como se ilustra a continuación.

II.1.- DISTRIBUCIÓN BINOMIAL[12]

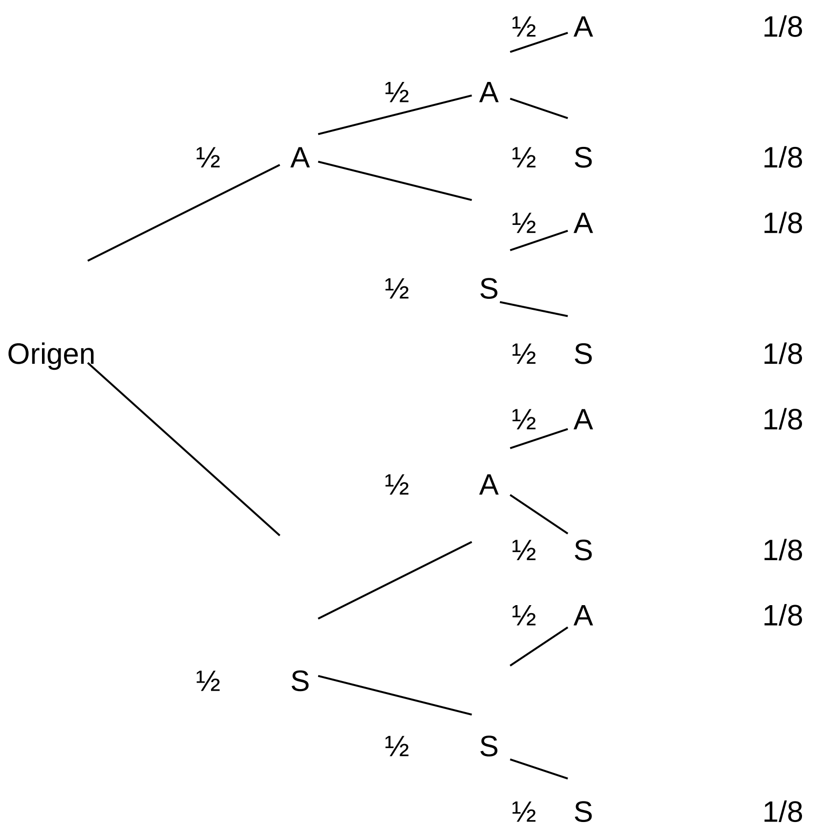
Es una familia de distribuciones con ciertas características en común. Una de sus principales características es que maneja datos discretos y no continuos. Se llama binomial porque se genera de la expansión binomial de **q+p**.

Puede obtenerse por medio de:

a) Diagrama de árbol.
b) La expansión binomial **q + p.**

Partiendo del Diagrama de árbol, en el caso del experimento consistente en el lanzamiento una vez de tres monedas al aire, estableciendo que p es A (águila) y q es S (sol) la distribución binomial gráficamente se generará así :

Probabilidad de resultados posibles



Agrupando los resultados anteriores en torno a una distribución probabilística tendremos:

No de águilas	Probabilidad
0	1 8
1	3 8
2	3 8

3	1 8

Para construir el diagrama de árbol se supone que los eventos son mutuamente excluyentes e independientes, lo cual quedó demostrado en la introducción a la probabilidad.

Ahora bien, para ilustrar la creación de la distribución binomial mediante la expansión del binomio, **(q + p)**. Para ello supongamos que una moneda ahora se lanza al aire dos veces y nos interesa obtener la probabilidad de que caigan "águilas". Los resultados posibles son 0, 1, 2 "águilas"; así mismo en el caso de una moneda no deforme, en cada lanzamiento la probabilidad de obtener águila (p) es 0.5 y la de sol es también 0.5 = q ; tal que **q + p** = 0.5 + 0.5 = 1. Luego la distribución binomial se obtiene de $\mathbf{(q + p)^n}$ donde n = 2 lanzamientos de la moneda. Así, con x representando águilas.

X	P(X)
0	0.25
1	0.50
2	0.25
	1.00

Sustituyendo las literales q y p hacemos: $(0.5+0.5)^2 = (0.5)^2 + 2(0.5)(0.5) + (0.5)^2$

$= 0.25 + 0.50 + 0.25 = 1.00$

P (0) = 0.25

P (1) = 0.50

P (2) = 0.25

Esto es una distribución binomial. Es probabilística porque muestra cada resultado posible con su probabilidad de ocurrencia asociada.

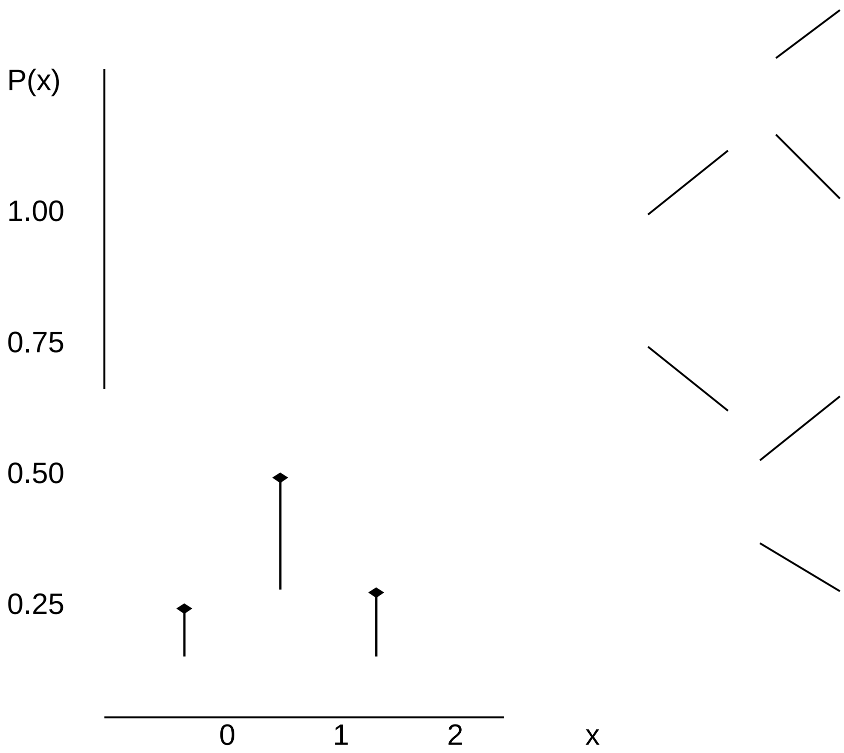
Gráficamente.

A ¼

½

A

½



Importante: Recordando que la probabilidad en su acepción objetiva se refiere a un proceso repetitivo, el cual genera productos que no son idénticos ni predecibles individualmente, pero que pueden describirse en términos de frecuencias relativas, estos procesos son llamados **estocásticos ó aleatorios**, y los resultados individuales se llaman eventos.

Un proceso estocásticos puede ser el lanzamiento de una moneda, el proceso de fabricación de ladrillos o la selección al azar de personas y el registro de su peso, estatura, ingreso o sexo.

En estos caso el proceso estocástico es el lanzamiento de la moneda, la fabricación de ladrillos ó el registro de las características de las personas, y lo que se observa (cara de la moneda, el peso de los ladrillos, el ingreso de las personas, etc.) es llamado variable estocástica, aleatoria o al azar.

De esta manera una distribución de probabilidad es una lista de todos los eventos (o valores de la variable aleatoria) que resulta de un proceso estocástico, y la probabilidad asociada de ocurrencia de cada uno de ellos.

Ahora bien, usando otras literales para evitar la rutina, si p: probabilidad de éxito de x y q: probabilidad de fracaso de x, los valores de p y q pueden variar pero su suma será **$q + p = 1$**.

Observaciones:

1.- El número de eventos en la secuencia o número de repeticiones se indica con el exponente del binomio. Así **$(q + p)$** es la expansión binomial que genera una distribución de probabilidad cuando se lanza una moneda.

Por consiguiente **$(q + p)^3$** es la expansión binomial que genera una distribución de probabilidad cuando se lanzan tres monedas, el término binomial a expandir será:

$$(q + p)^3 = q^3+3pq^2+3p^2q+p^3$$

Sustituyendo los valores de q y p, donde $q = 1/2 = p$; tendremos:

$$(q + p)^3 = (1/2)^3 + 3\,(1/2)\,(1/2)^2 + 3\,(1/2)^2\,(1/2) + (1/2)^3 =$$

$$= 1/8+3/8+3/8+1/8$$

Estos resultados son iguales a los obtenidos con el diagrama de árbol y corresponden a la probabilidad de obtener 0, 1, 2 ó 3 águilas en el lanzamiento de 3 monedas.

El primer término de la expansión indica la probabilidad de obtener cero águilas y tres soles, el segundo expresa la probabilidad de obtener un águila y dos soles y así sucesivamente. Luego los exponentes incluidos en cada término de la expansión binomial son útiles en la interpretación del significado de cada uno de los términos.

2.- Por otro lado, los coeficientes de cada término indican el número de formas en que se pueden obtener los resultados.

En resumen, la distribución binomial puede generarse de dos maneras:

1º.- Por el diagrama del árbol.

2º.- Por la expansión del binomio **$(q + p)^n$**

II.1.1 LA MEDIA ARITMÉTICA Y DESVIACIÓN ESTÁNDAR DE LA DISTRIBUCIÓN BINOMIAL

Se calculan con el procedimiento usual, solo que se usan probabilidades en lugar de frecuencias. En el caso de la media:

$$« = \frac{\sum Xp(X)}{\sum p(X)} \quad \text{en lugar de} \quad \overline{X} = \frac{\sum xf}{\sum f}$$

$$\pm = \sqrt{\frac{\sum (x - «)^2 p(x)}{\sum p(x)}} \quad \text{en lugar de} \quad \pm = \sqrt{\frac{\sum (x - \overline{x})^2 f}{\sum f}}$$

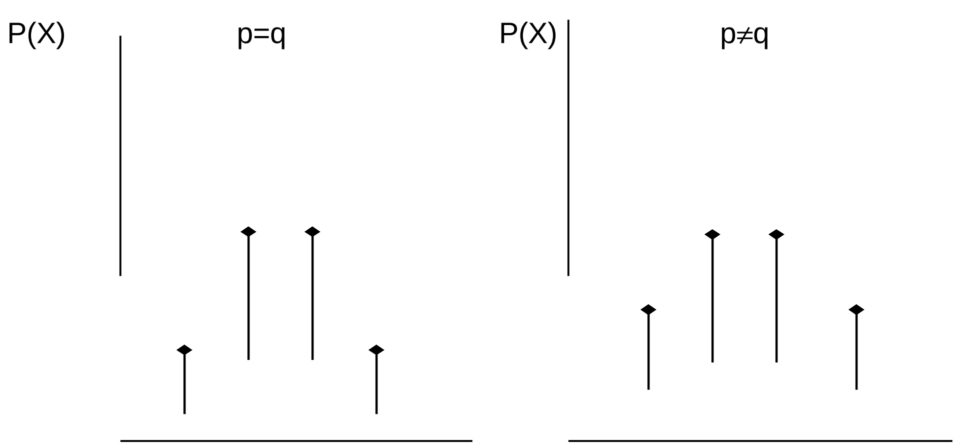
Como la suma de las probabilidades es igual a 1 los denominadores de las fórmulas se eliminan y queda:

$$« = \sum xp(x)$$

$$\pm = \sqrt{\sum (x - «)^2 p(x)}$$

La distribución binomial es simétrica cuando p = q = 1/2; y asimétrica cuando p es diferente de q.

Gráficamente:



0 1 2 3 x 0 1 2 3 x

Ejemplo 1

Si el 50% de los hombres empleados en la Cía. Nestle son casados y si tomamos una muestra aleatoria de dos hombres, ¿Cuál es la probabilidad de que la muestra contenga 2, 1 ó 0 hombres casados?

p = 1/2 = q

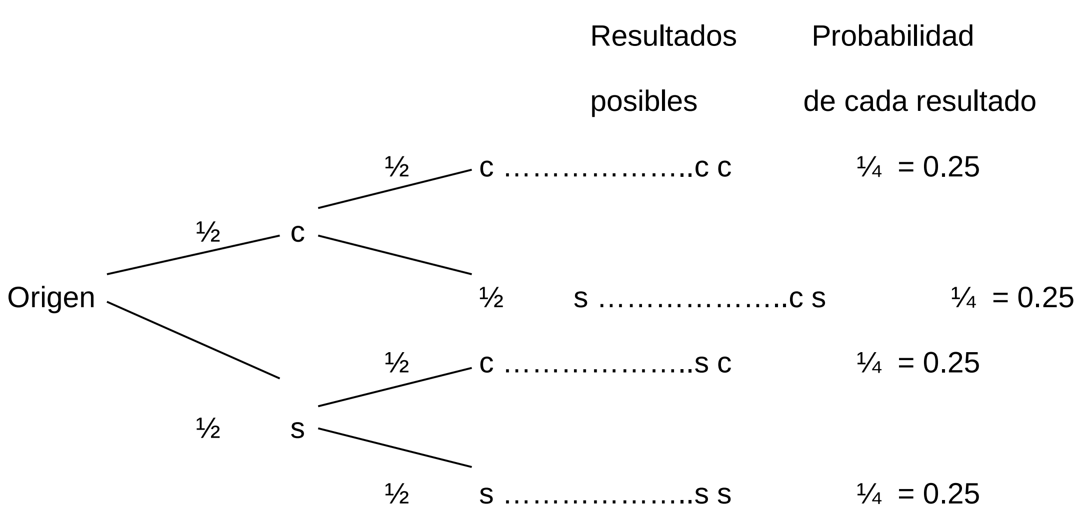
p : probabilidad de que los hombres sean casados.

q : probabilidad de que no lo sean.

c = casado.

s = soltero.

En este caso usando El diagrama de árbol, la distribución binomial será:



1.00

Agrupando los resultados en una tabla de frecuencias (probabilidades) relativas, tenemos:

X	P(X)
0	0.25
1	0.50
2	0.25
	1.00

Este mismo resultado puede obtenerse con la expansión del binomio

$\mathbf{(q + p)^2}$

$\mathbf{(q + p)^2 = q^2+2pq+p^2}$

$= (1/2)^2+2(1/2)(1/2)+(1/2)^2$

$= 1/4+2(1/4)+1/4$

$= .25+.50+.25=1$

CÁLCULO DE LA MEDIA ARITMÉTICA Y DESVIACIÓN ESTÁNDAR DE LA DISTRIBUCIÓN BINOMIAL.

X	P(X)	XP(X)	x-μ	$(x-\mu)^2$	$(x-\mu)^2P(X)$
0	0.25	0.00	-1.00	1.00	0.25
1	0.50	0.50	0.00	0.00	0.00
2	0.25	0.50	1.00	1.00	0.25
	1.00	1.00	0.00		0.50

Se calcula con el procedimiento usual, solo que se usan probabilidades en lugar de frecuencias. En el caso de la media, μ:

$$« = \frac{\sum xp(x)}{\sum p(x)} \text{ en lugar de } \overline{X} = \frac{\sum xf}{\sum f}$$

μ=1

$$\pm = \sqrt{\frac{\sum (x-«)^2 p(x)}{\sum p(x)}} = \sqrt{0.50} = 0.71 \text{ en lugar de } \pm = \sqrt{\frac{\sum (x-\overline{x})^2 f}{\sum f}}$$

Estos resultados μ y σ se obtiene más fácilmente con:

$\mu =$ n p ; y $\pm = \sqrt{npq}$

donde n = número de veces que se realiza el experimento o tamaño de la muestra:

Si p = 1/2 y n = 2 ; $\mu = 2(1/2) = 1$

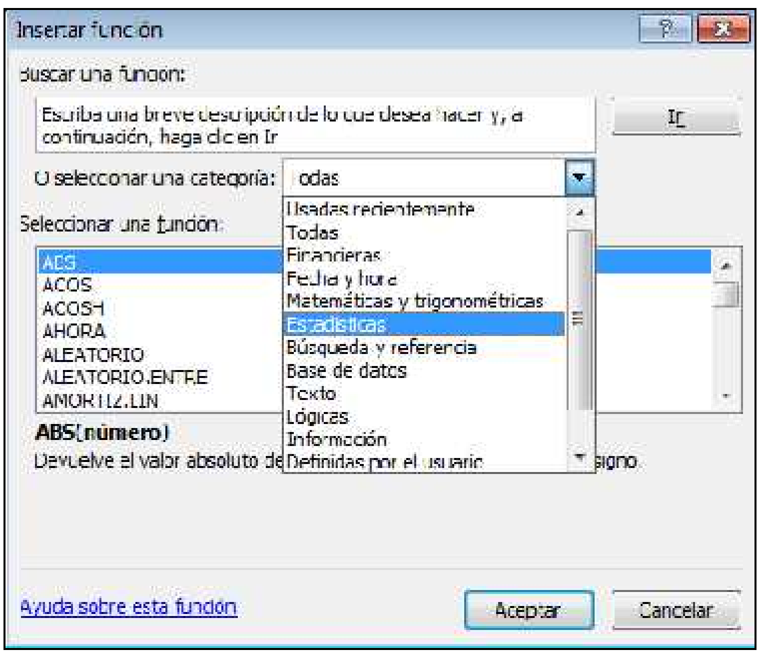
$$\pm = \sqrt{2(1/2)(1/2)} = 0.71$$

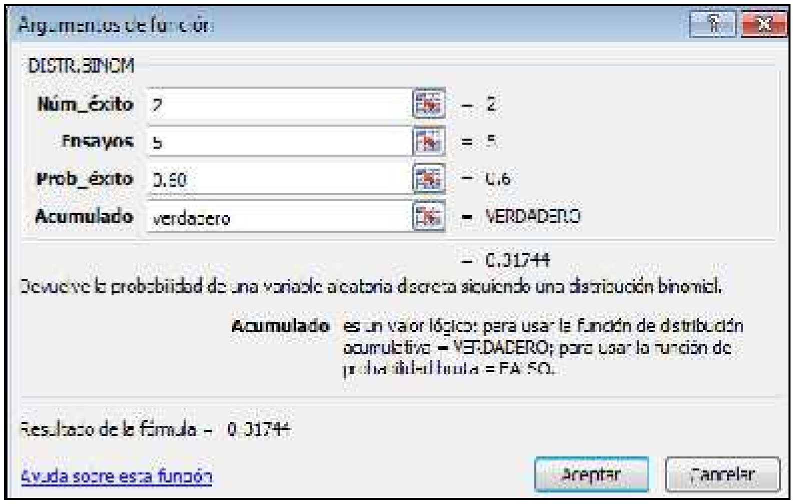
II.1.1.1. GENERACIÓN DE LA DISTRIBUCIÓN BINOMIAL CON EXCEL.

Con las referencias anteriores ahora suponga que n=5; p=0.60 y que x=2

PASOS:

- Colocamos el cursor en **fx /** click **/** insertar una función/ *estadísticas /* click/ **DIST.BINOM**/ click/ aceptar/argumentos de función: en *número de éxitos*=2; ensayos= 5; en *prob_éxito* =0.60 y en *acumulado*: **verdadero**/ aceptar

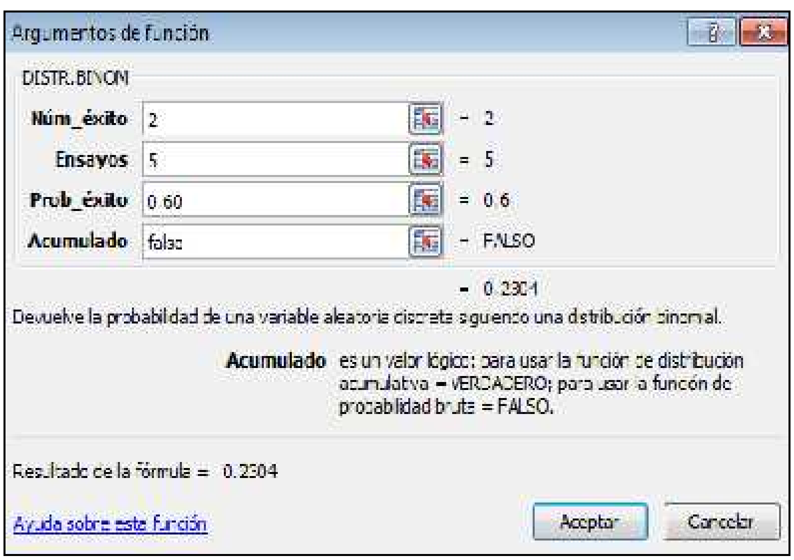




Esta será la probabilidad de que p(x sea menor o igual a 2)=0.31744, que es la probabilidad acumulada de x=0+x=1+x=2.

- Si queremos calcular exactamente la probabilidad de ocurrencia de 2, entonces en la celda de *acumulado* escribimos **falso**/aceptar y obtenemos p(x=2)=0.2304.

  Luego entonces con este último procedimiento obtenemos la probabilidad de cada uno de los n+1 resultados posibles cuando el ensayo o experimento se hace 5 veces.



II.1.1.2 LA DISTRIBUCIÓN NORMAL COMO LÍMITE DE LA BINOMIAL

Hemos visto que la distribución binomial es discreta porque sus términos son enteros. El polígono de frecuencias no tiene otro significado que el de ilustrar su simetría o asimetría por lo que no se pueden interpolar sus puntos al no ser fraccionables sus valores.

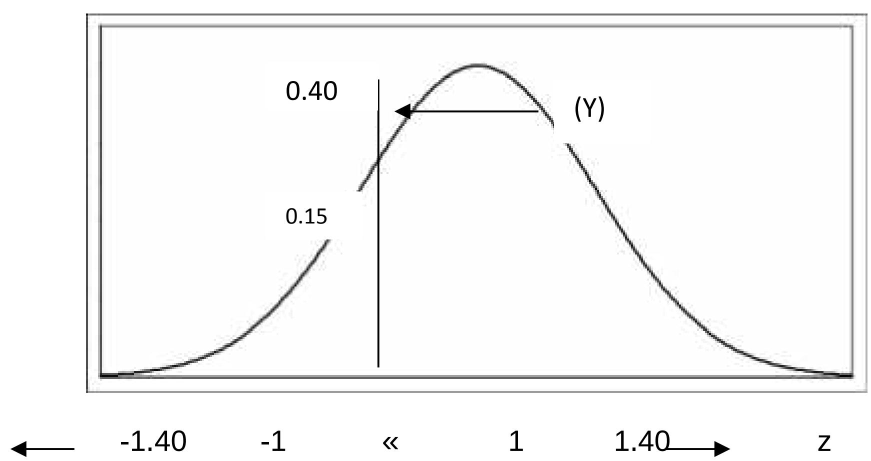
Sin embargo cuando n crece también lo hace el número de resultados posibles, tal que el polígono de frecuencias se aproxima a una curva suave y acampanada que corresponde a la distribución normal. La distribución normal fue derivada de la estandarización de la binomial usando:

la variable $Z = \frac{(x - «)}{±}$ que es igual a $Z = \frac{(x - np)}{\sqrt{npq}}$ y con n creciendo sin límite.

Si usamos el ejemplo anterior para ilustrar esta transformación y usando las nuevas literales, tendremos:

X	P(X)	x-μ	$Z = \frac{x - «}{±}$	Área bajo la curva	(y) Ordenada
0	0.25	-1.00	-1.40	-0.41924	0.14973
1	0.50	0.00	0.00	0	0.39894
2	0.25	1.00	1.40	0.41924	0.14973

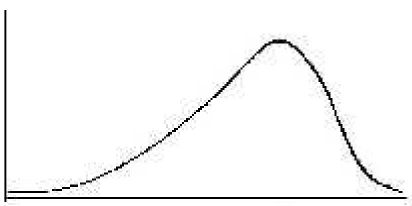
μ = n p = 1; σ = 0.71



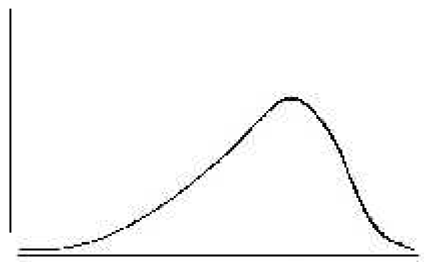
La normal es simétrica aún cuando p es diferente de **q** tal que la binomial con **p** diferente de **q** pero con n creciendo tiende a ser normal o simétrica.

Ejemplo :

n=2



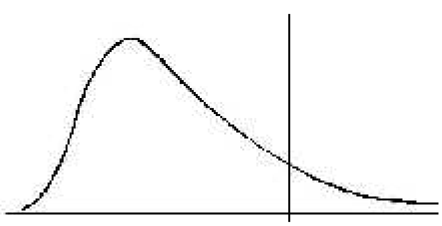
n=20



0 1 2 x x

P(X)

n=200



x

La distribución binomial también se le llama de Bernoulli, porque fue quien la desarrolló.

II.2.- DISTRIBUCIÓN HIPERGEOMÉTRICA

A la distribución de probabilidad determinada a partir de una **distribución finita** de datos discretos se le llama distribución hipergeométrica, En otras palabras, cuando la población estadística es pequeña se conoce su tamaño y por consiguiente es finita. Si de ella seleccionamos una muestra, n, que sea mayor a 5% de la población, N, y si se usa un muestreo con reemplazo, en ese caso se aconseja utilizar la distribución hipergeométrica en lugar de la binomial para calcular la probabilidad de un número específico de éxitos (x), cuya probabilidad de cada uno de ellos no es la misma debido al reemplazo de los elementos en la selección de la muestra de la población que es pequeña y por ende es finita.

Indudablemente que si la población fuera grande y la selección de la muestra fuera con reemplazo *se podría asignar la misma probabilidad* a cada uno de los elementos para ser incluidos en la muestra.

Para su cálculo se parte de las fórmulas de la binomial obtenida con la formula de las **combinaciones**:

$$\begin{bmatrix} n \\ r \end{bmatrix} = \frac{n!}{r!(n-r)!} = {}_n C_r$$

En este caso tenemos que:

$$\pm = \sqrt{npq} * \sqrt{\frac{N-n}{N-1}}$$

Conociéndose $\sqrt{\frac{N-n}{N-1}}$ con el nombre del multiplicador ó corrector finito, el cual es útil porque ayuda a mejorar el valor de r.

Ejemplo 1.-

N = Universo = 200 automóviles

n_1 = Automóviles americanos = 120

n_2 = Automóviles europeos = 80

r = Tamaño de la muestra = 20

¿Cuál es la probabilidad de que x = 8 sean americanos?. Recordando que habrá $\begin{bmatrix} n_1 \\ x \end{bmatrix}$ maneras diferentes de obtener 8 automóviles americanos, entonces r - x: será el número de automóviles europeos tal que $\begin{bmatrix} n_2 \\ r-x \end{bmatrix}$ maneras diferentes de obtener 12 automóviles europeos.

Luego la probabilidad de obtener 8 automóviles americanos y 12 europeos será:

$$\frac{\begin{bmatrix} n_1 \\ x \end{bmatrix}\begin{bmatrix} n_2 \\ r-x \end{bmatrix}}{\begin{bmatrix} N \\ r \end{bmatrix}} = \frac{\begin{bmatrix} 120 \\ 8 \end{bmatrix}\begin{bmatrix} 80 \\ 12 \end{bmatrix}}{\begin{bmatrix} 200 \\ 20 \end{bmatrix}}$$

La distribución hipergeométrica se genera para todos los éxitos (X).

X		P (X)
Autos Americanos		
0	.	$P(x=0)=\dfrac{\begin{bmatrix}120\\0\end{bmatrix}\begin{bmatrix}80\\20\end{bmatrix}}{\begin{bmatrix}200\\20\end{bmatrix}}$
1	.	
2	.	
3	.	
.		.
.		.
.		.
8	.	$P(x=8)=\dfrac{\begin{bmatrix}120\\8\end{bmatrix}\begin{bmatrix}80\\12\end{bmatrix}}{\begin{bmatrix}200\\20\end{bmatrix}}$
.		.
.		.
20	.	$P(x=20)=\dfrac{\begin{bmatrix}120\\20\end{bmatrix}\begin{bmatrix}80\\0\end{bmatrix}}{\begin{bmatrix}200\\20\end{bmatrix}}$
Suma		1.00

Ejemplo 2

N = 10 personas

n_1 = 6 hombres

n_2 = 4 mujeres

r = 5

¿Cuál es la probabilidad de obtener hombres en una muestra de 5?

Número de hombres Combinaciones P(X)

(X)

0	$\dfrac{\begin{bmatrix}6\\0\end{bmatrix}\begin{bmatrix}4\\4\end{bmatrix}}{\begin{bmatrix}10\\5\end{bmatrix}} = \dfrac{1}{252} =$	0.0000
1	$\dfrac{\begin{bmatrix}6\\1\end{bmatrix}\begin{bmatrix}4\\4\end{bmatrix}}{\begin{bmatrix}10\\5\end{bmatrix}} = \dfrac{6(1)}{252} = \dfrac{6}{252} =$	0.0238
2	$\dfrac{\begin{bmatrix}6\\2\end{bmatrix}\begin{bmatrix}4\\3\end{bmatrix}}{\begin{bmatrix}10\\5\end{bmatrix}} = \dfrac{15(4)}{252} = \dfrac{60}{252} =$	0.2380
3	$\dfrac{\begin{bmatrix}6\\3\end{bmatrix}\begin{bmatrix}4\\2\end{bmatrix}}{\begin{bmatrix}10\\5\end{bmatrix}} = \dfrac{20(6)}{252} = \dfrac{120}{252} =$	0.4761
4	$\dfrac{\begin{bmatrix}6\\4\end{bmatrix}\begin{bmatrix}4\\1\end{bmatrix}}{\begin{bmatrix}10\\5\end{bmatrix}} = \dfrac{15(4)}{252} = \dfrac{60}{252} =$	0.2380
5	$\dfrac{\begin{bmatrix}6\\5\end{bmatrix}\begin{bmatrix}4\\0\end{bmatrix}}{\begin{bmatrix}10\\5\end{bmatrix}} = \dfrac{6(1)}{252} = \dfrac{6}{252} =$	0.0238
		.9757 $\cong$ 1.000

II.2.1 SU MEDIA ARITMÉTICA Y DESVIACIÓN ESTÁNDAR

Calcular la μ y la σ de la hipergeométrica con μ = np = Σ XP(X) y

$$\pm = \sqrt{\boldsymbol{npq}} * \sqrt{\frac{\boldsymbol{N-n}}{\boldsymbol{N}-1}} = \sqrt{\sum (\boldsymbol{x} - \text{«})^2\, \boldsymbol{p}(\boldsymbol{x})}$$

X	P(X)	XP(X)	x-μ	$(x\text{-}\mu)^2$	$(x\text{-}\mu)^2P(X)$
0	0.0000	0.000	-3	9	0.000
1	0.0238	0.024	-2	4	0.096
2	0.2380	0.476	-1	1	0.238
3	0.4761	1.428	0	0	0.000
4	0.2380	0.952	1	1	0.238
5	0.0238	0.120	2	4	0.096
		3.000			0.668

μ = Σ X P(X)

μ = 3

También se obtiene el mismo resultado con :

μ = n p

μ = 5(0.6) = 3

ya que p = 0.6 = probabilidad de obtener "hombre" en una selección simple o proporción de hombres en la población. Por su parte la desviación estándar será:

$$\pm = \sqrt{\sum (\boldsymbol{x} - \text{«})^2\, \boldsymbol{p}(\boldsymbol{x})} = \sqrt{0.668} = 0.81$$

Como en el caso de la media,   también se obtiene de :

$$\pm = \sqrt{\boldsymbol{npq}} * \sqrt{\frac{\boldsymbol{N}-\boldsymbol{n}}{\boldsymbol{N}-1}}$$

$$\pm = \sqrt{5(0.6)(0.4)} * \sqrt{\frac{10-5}{10-1}} = \sqrt{1.20} * \sqrt{0.55} = 0.81$$

El profesor Lind (et al 2005) comenta que esta distribución *debe usarse cuando*: a) la probabilidad de ocurrencia de cada evento (1/n) ya no sea la misma, como sucede en poblaciones pequeñas finitas de las que se extrae la muestra (n) usando el *muestreo si reemplazo*, dado que al no ser reemplazado el resultado posible que apareció en la primera selección de la muestra, en la segunda selección se dispone de menos resultados en el espacio muestral para ser incluidos en la muestra, y por ende la probabilidad de ser incluido cada resultado posible en la muestra ahora es 1/n-1; en una tercera selección, al quedar menos datos para ser incluidos en la muestra, la probabilidad de cada uno de ellos de ser incluidos en la muestra ahora es 1/n-2, etc, ; b) cuando n sea mayor al 5% de N (población estadística).

Por lo ilustrativo, juzgué conveniente mostrar su ejemplo, cuyo planteamiento es : Una empresa tiene 50 empleados (N) de los cuales 4 son sindicalizados y 10 no lo son. Con esas referencias, se tomará una muestra de 5 empleados para que participen en las negociaciones del nuevo contrato de trabajo que regirá sus relaciones de trabajo durante el año próximo. Lind se pregunta, ¿Cuál es la probabilidad de que 4 de los trabajadores sean sindicalizados?

Para contestar usa la fórmula de las combinaciones arriba descrita estableciendo que X representa a los trabajadores sindicalizados y desarrolla las probabilidades de que ocurra cada una de las X, es decir P(X), arribando a la siguiente tabla *que representa la distribución hipergeométrica:*

X	P(X)
0	0.0000
1	0.004
2	0.044
3	0.220
4	0.431
5	0.311

Total	1.0000

Luego la P(X=4) será 0.431 como la P(X=5)=0.311, etc.

II.3 DISTRIBUCIÓN DE POISSON

La distribución de Poisson también es discreta y forma parte de la familia Bernoulli.

La distribución binomial se aproxima a la normal cuando **n** crece aunque **q** sea diferente de **p**. Sin embargo cuando p es pequeña la aproximación de la binomial a la normal no es satisfactoria, por lo que la distribución de Poisson deberá usarse como una aproximación.

En este caso la probabilidad de **x** eventos en **n** pruebas, cuando **p** es la probabilidad de que suceda dicho evento en una prueba simple viene dada por:

$$P(X) = e^{-np} * \frac{(np)^x}{X!}$$

Si **np**=**m** entonces $P(X) = e^{-m} * \frac{(m)^x}{X!}$

e es la base de los logaritmos naturales = 2.71828

Como la binomial, la media de la distribución de Poisson es **np** = **m**, pero su varianza es **<u>m</u>** porque si:

$\sigma^2 = n\,p\,q$ y si $q \cong 1$, entonces σ^2, $= n\,p = m$

Ejemplo:

El gimnasio Gumesindo Brown del D.F. pide un aparato de ejercicios a Monterrey; este es enviado con 200 tuercas para ser armado aún cuando sólo requiere 198. Las dos tuercas adicionales son incluidas como reserva para que en caso de que salieran defectuosas algunas se pudieran substituir con las dos de repuesto. Las tuercas son hechas por una máquina automática que produce tuercas defectuosas con una probabilidad de 0.01.

¿Cuál es la probabilidad de que el comprador no tenga suficientes

tuercas para armar el aparato?

p = 0.01

m = n p = 200 (0.01) = 2

n = número total de tuercas = 200

$$P(X) = e^{-m} * \frac{(m)^x}{X!}$$

$e^{-2} = \frac{1}{(2.71828)^2} = 0.13534$ por lo tanto e^{-m} = 0.13534

X	P(X)	
0	0.1353	P(0)= 0.13534*$(2)^0$/0! = 0.13534
1	0.2707	P(1)= 0.13534*$(2)^1$/1! = 0.2707
2	0.2767	P(2)= 0.13534*$(2)^2$/2! = <u>0.2707</u>
		0.6767

Luego si P(x > 2) = 1.000 - 0.6767

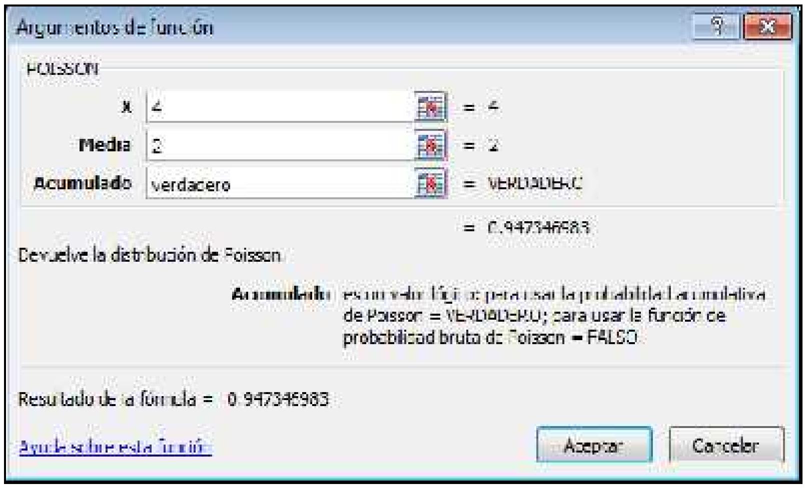
Entonces p(x > 2) = 0.3232

GENERACIÓN DE LA DISTRIBUCIÓN DE POISSON CON EXCEL

- **DISTRIBUCIÓN DE POISSON: (CIRO MARTINEZ: 2005:93)**

Fórmula: $P_{(x)} = \frac{\lambda^X e^{-\lambda}}{x!}$

- Colocamos el cursor en **fx/** click **/** insertar una función/ *estadísticas /* clic/ **BINOMIAL**/ click/ aceptar/argumentos de función: x=4; media (lamda)= 2; en y en *acumulado*: **verdadero**/ aceptar

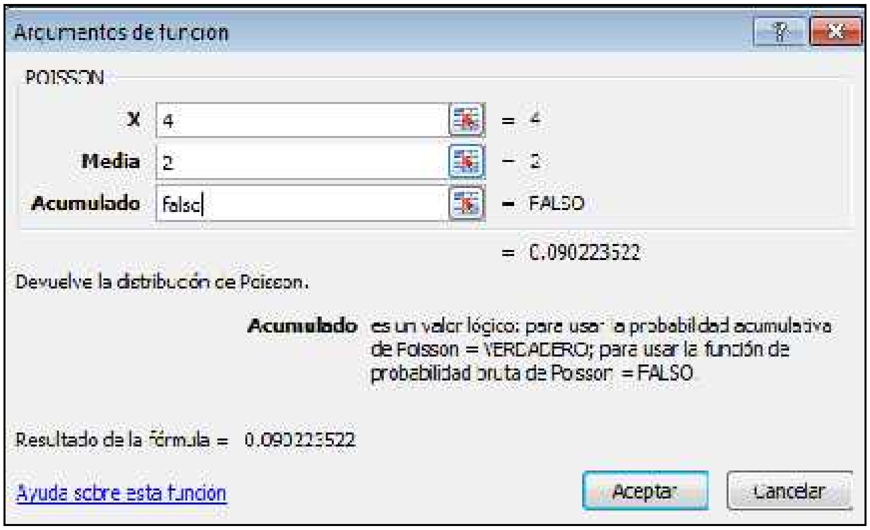


$$P_{(x \leq 4)} = P_{(x=0)} + P_{(x=1)} + P_{(x=2)} + P_{(x=3)} + P_{(x=4)} =$$

$$P_{(x=4)} = \frac{2^0 e^{-2}}{0!} + \frac{2^1 e^{-2}}{1!} + \frac{2^2 e^{-2}}{2!} + \frac{2^3 e^{-2}}{3!} + \frac{2^4 e^{-2}}{4!} = 0{,}947346$$

Tal como lo podemos observar, en la parte inferior de la figura, donde dice *resultado de la fórmula*, aparece: 0,947346987

- Sí, por el contrario, digitamos la palabra FALSO, se obtiene la probabilidad de que ocurra exactamente, por ejemplo, en nuestro caso CUATRO: x = 4



$$P_{(x\,=\,4)} = \frac{\lambda^{X}\, e^{-\lambda}}{x!} = \frac{2^{4}\, e^{-2}}{4!} = 0{,}090225522$$

II.4.- DISTRIBUCIÓN NORMAL

La ecuación para obtener sus ordenadas, dados ciertos valores de X expresados en las abscisas en unidades de desviación estándar, es:

F(x)= 1/ $\sqrt{2®±^{2}} \; * \; e^{\left[-1/2±^{2}(x-«)^{2}\right]}$

donde:

x = cualquier valor de la variable aleatoria continua;

« = media de la variable aleatoria normal

± = desviación estándar de la variable aleatoria normal;

$e = 2.71828$...base del logaritmo natural

$® = 3.1416$,

dándole valores podemos construir la curva normal

Características:

Tiene forma de campana; es simétrica respecto a su media; es asíntota al eje de las x o sea que nunca atraviesa el eje de las x.

El área bajo la curva normal es igual a 1 dado que representa la suma de las probabilidades de todos los resultados posibles de un experimento.

En la vida real hay distribuciones de datos con medias iguales y desviaciones estándar diferentes o con medias diferentes y desviaciones estándar iguales.

Para uniformarlas o reducirlas a un patrón único (Hayashi et al, 1974), hacemos un cambio de variable , que designamos con $Z = \frac{x - «}{±}$ y llamamos **variable normal estándar**, misma que al ser una desviación de los términos x con respecto a su media en forma estandarizada otros autores la llaman **Desviación normal estandarizada** , la cual que tiene $« = 0\, y ± = 1$, demostración:

su promedio será $\bar{Z} = \frac{\sum \frac{x - «}{±}}{N}$, como ± es una constante la podemos

sacar de la sumatoria $\bar{Z} = \frac{\frac{1 \sum (x - «)}{±}}{N}$

la suma de la diferencia $x - « = 0$, luego

$\bar{Z} = \frac{\frac{1[0]}{±}}{N}$ así tenemos $\bar{Z} = \frac{0}{N} = 0$, lo que queda demostrado

Ahora bien, demostrar que $±\ z = 1$

$$\pm_z = \sqrt{\frac{\sum\left(\frac{x-\text{«}}{\pm}-\bar{z}\right)^2}{N}} = \sqrt{\frac{\sum\left(\frac{x-\text{«}}{\pm}-0\right)^2}{N}}$$

$$\pm_z = \sqrt{\frac{\sum\left(\frac{x-\text{«}}{\pm}\right)^2}{N}} = \sqrt{\frac{\sum\frac{(x-\text{«})^2}{\pm^2}}{N}} = \sqrt{\frac{\sum\frac{1}{\pm^2}(x-\text{«})^2}{N}}$$

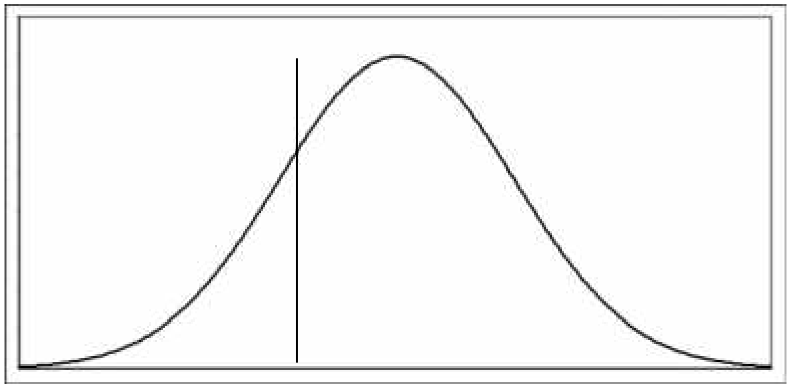
Al ser $\pm^2$ una constante la podemos sacar de la sumatoria

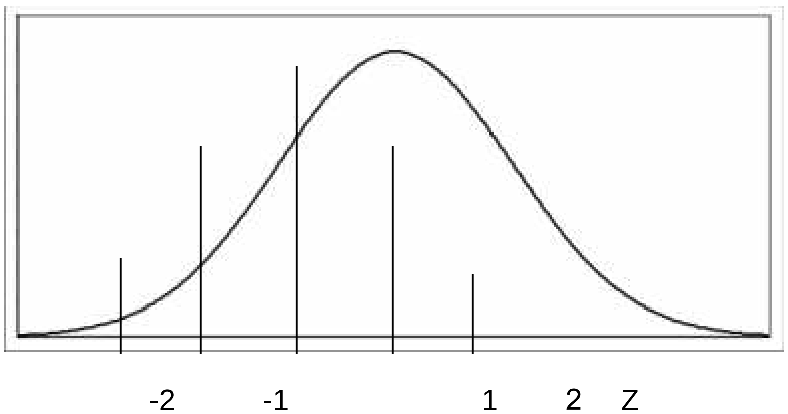
$$\pm_z = \sqrt{\left(\frac{1}{\pm^2}\right)\frac{\sum(x-\text{«})^2}{N}} = \frac{1}{\pm}\sqrt{\frac{\sum(x-\text{«})^2}{N}}$$

Sabemos que $\pm = \sqrt{\frac{\sum(x-\text{«})^2}{N}}$

Luego $\pm_z = \frac{1}{\pm} * \pm = \frac{\pm}{\pm} = 1$

Al contar con la variable Z, que expresas los valores originales expresados en términos de su desviación estándar, ahora ya podemos utilizar los valores de Z que están en el apéndice A para analizar e interpretar cualquier fenómeno económico en términos de sus valores estandarizados. Esta situación aumenta la capacidad de estudio o caracterización del citado fenómeno económico al poder ahora el investigador calcular valores esperados, determinar limites de confianza dentro de los cuales pueda ocurrir un cierto valor del mismo, hacer estimaciones e inclusive probar ciertas hipótesis de interés para el investigador, como se muestra a continuación sabiendo que la normal es una distribución teórica como la binomial, poisson e hipergeométrica, pero con datos continuos que nos ayuda a hacer análisis. Su figura o forma es:





Pero ¿cómo se obtiene, cómo se gráfica? ¿de dónde provienen los valores estandarizados de Z?

Ejemplo 1 que ilustra cómo se construye la curva normal

Supongamos que el salario promedio de 15 000 obreros es de 900 pesos con una desviación estándar de 150 pesos. Construya usted la curva normal con intervalos de 1/2 σ, a partir de μ hasta tres veces sigma.

N = 15,000

μ =$ 900.00

σ =$ 150.00

a) Construya la curva normal.

Comentarios: Los valores de las coordenadas para trazarla vienen en la primera (valores de Z) y tercera columna (valores de las **ordenadas para una población infinita, que fueron calculadas con la ecuación descrita al inicio del tema)del Apéndice A;** basta darle valores a z y encontrar sus ordenadas correspondientes para graficar en los cuadrantes I y II (ya que la curva es simétrica) del sistema de ejes cartesianos, un número suficiente de puntos, que al unirlos verificaremos que obtenemos una figura idéntica a una campana.

Sin embargo, cuando la población es finita, como es el caso de este ejemplo, porque conocemos N, y sabiendo que Z es la abscisa de cada uno de los puntos que nos van a permitir construir la curva, **se debe calcular o acotar, su ordenada** correspondiente a partir de los valores presentados en la tercera columna del Apéndice A, con la siguiente fórmula:

Yx = N / σ * f(Z)

Donde $Z = \frac{X_i - \text{«}}{\pm}$ del Apéndice A.

Z ; es el valor de la abscisa o dicho en otras palabras, es el valor expresado en unidades de desviación estándar, de cada uno de los valores originales denotados con los símbolos X_i

Valores originales	Inicio de la conversión a unidades Z	Obtención de Z	Ordenadas para cada valor de Z en una población infinita	Determinación de las ordenadas para esta población finita
Xi	Xi-µ	$Z = \frac{X_i - \text{«}}{\pm}$	f(Z)	Yx
900	0	0.00	0.39894	39.89
975	75	0.50	0.35207	35.2
1050	150	1.00	0.24197	24.19

1125	225	1.50	0.12952	12.95
1200	300	2.00	0.054	5.4
1275	375	2.50	0.175	1.75
1350	450	3.00	0.0044	0.44

Tabulaciones:

Yx = 15,000 /150 * f(Z)

Yx = 100 (f(Z))

La columna f(Z) se encuentra en las tablas del apéndices A, buscando primero $Z = \frac{X_i - \text{«}}{\pm}$ que lo tenemos ya en las columnas arriba.

Por ejemplo si Z = 0 lo buscamos en la primera columna de las tablas apéndice A, una vez encontrado nos pasamos a buscar f(Z), que estará en la columna tres de las tablas.

Ejemplo 2

El tiempo de duración de 5 000 pilas para tomar fotografías producidas por una compañía están normalmente distribuidas con media igual a 800 minutos y σ = 40 minutos.

a) Construya gráficamente la curva normal correspondiente con intervalos de 1/2 de σ hasta 3 veces.
b) ¿Cuántas pilas duran entre 780 y 820 minutos?
c) ¿Cuál es la probabilidad de que al seleccionar una pila esta dure cuando menos 750 minutos?

N = 5 000; μ = 800; σ = 40

a)

Xi	Xi-µ	$Z = \frac{X_i - «}{±}$	f(Z)	Yx
800	0	0.00	0.39894	49.86
820	20	0.50	0.35207	44
840	40	1.00	0.24197	30.24
860	60	1.50	0.12952	16.19
880	80	2.00	0.05399	6.74
900	100	2.50	0.01753	2.19
920	120	3.00	0.00443	0.55

Yx = N / σ * f(Z)

Yx = 5000 / 40 = 125

Yx = 125 * f(Z)

Ahora bien, para contestar el inciso b) tenemos que determinar Z_1 y Z_2 con

Z = Xi - µ / σ

Z_1 = 780 - 800 / 40 = -0.5 unidades de desviación estándar, cuya área es 0.1915

Z_2 = 820 - 800 / 40 = 0.5 unidades de desviación estándar, cuya área es 0.1915

Luego entonces,

P{(X)} = El área de Z_1 = -0.5 a Z_0 + el área de Z_0 a Z_2 = 0.5

P(X) = 0.1915 + 0.1915 =0.383

Para saber cuántas pilas son: 5000(0.383) = 1915 pilas

Aprovechamos para reiterar que la Esperanza matemática E(X) = n p = 100 (0.383) = 38.3%

Para contestar la siguiente pregunta, hacemos:

c) Partiendo de $Z = \frac{X_i - «}{±}$

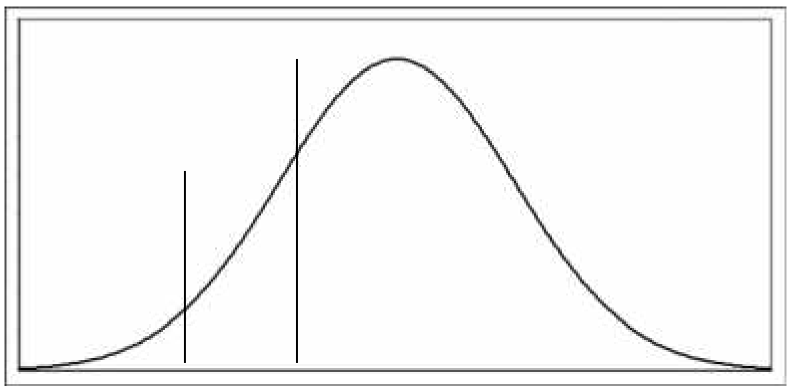
Sabemos que $Z = Xi - \mu / \sigma$

$Z_1 = 750 - 800 / 40 = -1.25$ unidades de desviación estándar

P{(X)}: el área de $Z_1 = -1.25$ luego el área correspondiente será:

de 0.39435.

$P(X \geq 750) = 0.39435 + 0.5000 = 0.89435$ ó 89.435 %.



750 800

Aproximación de la normal a la binomial.

En una distribución binomial cuando n es grande, es difícil calcular los resultados; en este caso se aplica la aproximación normal a la binomial, digamos cuando $n > 25$. Ejemplo:

Problema I. La compañía de seguros GNP realizó un estudio entre los dueños de bienes asegurados, con objeto de conocer cuántos recuperaban del total que reclamaban por robo. Halló que no se recuperaba el 80% de los hurtos reportados a GNP. Con ese referente:

1. Si se reportaron 200 robos, ¿Cuál es la probabilidad de que no se recuperen los bienes hurtados **en más** de 170 de los actos de latrocinio?

Con $p = 0.8$ que indica que no recuperaron y si $n = 200$ entonces:

$\mu = np = 200(0.8) = 160$; y

$\sigma^2 = npq = 200(0.8)(0.2) = 32$

luego $\sigma = \sqrt{32} = 5.66$

Así, $Z = \frac{170-160}{5.66} = 1.68$ Cuya área es 0.4535 por tanto

$P(x > 170) = 0.0465$.

2. Si se reportaron 200 robos, ¿Cuál es la probabilidad de que no se recuperen los bienes robados **en más** de 150 de los delitos?

Si $\mu = 160$ y $\sigma = 5.66$ entonces

$Z = \frac{150-160}{5.66} = -1.68$ por tanto su área es 0.4535, luego

$P(x > 150) = 0.5000 + 0.4535 = 0.9535$

Problema II. Suponga que "x" tiene una distribución probabilística binomial, con n = 50 y p = 0.25 calcule:

1. Su media y su desviación estándar.

Con n = 50 y p = 0.25 tenemos $\mu = np = 50(0.25) = 12.5$ y

$\sigma^2 = npq = 50(0.25)(0.75) = 9.375$; tal que $\sigma = \sqrt{9.375} = 3.0619$

2. La probabilidad de que x > 25.

P(x > 25) = 0.2119 porque $Z = \frac{25-12.5}{3.0619} = 4.08$, su área es 0.2881, por lo que

P(x > 25) = 0.5000 – 0.2881 = 0.2119

3. La probabilidad de que x < 10.

P(x < 10) = 0.2119 dado que $Z = \frac{10-12.5}{3.0619} = –0.81$ cuya área es 0.2881, por tanto P(x < 10) = 0.5000 – 0.2881 = 0.2119

4. La probabilidad de que x esté entre 10 y 25 inclusive.

P(10 ≤ x ≤ 25) = 0.2881 + 0.5000 = 0.7881

Problema III. La SHCP en 2007 al hacer la devolución de impuestos federales, detectó que se cometió un error en el 10% de las devoluciones. Suponga que en 2008 se mantiene en ese porcentaje y se elaboraron 60 devoluciones de impuestos, ¿Cuál es la probabilidad de que cometa:

1. Más de 7 errores.

Si µ = np = 60(0.10) =6; σ^2 = npq = 60(0.9)(0.1) = 5.4; σ = $\sqrt{5.4}$ = 2.32, luego

$Z = \frac{7-6}{2.32} = \frac{1}{2.32} = 0.43$ cuya área es 0.1664, por consiguiente:

P(x > 7) = 0.5000 – 0.1664 = 0.3336

2. Por lo menos 9 errores.

Decimos que $Z = \frac{9-6}{2.32} = \frac{3}{2.32} = 1.29$ que tiene un área de 0.4015, de manera que la P(x ≥ 9) = 0.5000 – 0.4015 = 0.0985

Problema IV. Un estudio realizado por el Gimnasio “γ” indica que 30% de sus socios tienen un sobrepeso significativo. Hizo una promoción y logró inscribir 500 nuevos socios. ¿Cuál es la probabilidad de que:

1. Más de 175 de los nuevos socios tengan un sobrepeso considerable.

Sabemos que μ = np = 500(0.3) = 150 y que σ^2 = npq = (500)(0.3)(0.7) = 105.0, por lo que σ = $\sqrt{105} = 10.24$; así $Z = \frac{175-150}{10.24} = \frac{15}{10.24} = 1.47$ cuya área es 0.4292, tal que P(x > 175) = 0.5000 – 0.4292 = 0.0708

2. Más de 140 de ellos tengan un sobrepeso considerable.

Calculamos $Z = \frac{140-150}{10.24} = \frac{-10}{10.24} = -0.97$ y si su área es 0.3340, tenemos que

P(x > 140) = 0.5000 + 0.3340 = 0.8340

II.5.- PRÁCTICA

Ejercicio Nº 1.

Nombre ____ No de Cta.______________________ Grupo____________

Problema 1.-Con r = reprobado y nr = no reprobado, si sabemos que la probabilidad (p) de reprobar en el examen de Estadística es de 0.4, (es decir r), cuando tomamos una muestra de aleatoria de 4 alumnos, obtenga:

1.- La distribución probabilística correspondiente con el método de la expansión del binomio;

2.- Interprete los coeficientes y los exponentes de cada uno de sus términos;

3.- su media aritmética y desviación estándar, con cualesquiera de los métodos conocidos;

4.- Su gráfica e indique si es una distribución simétrica, ¿por qué?

5.- Transforme los datos discretos en continuos.

6.- ¿Cuál es la probabilidad de que ningún alumno repruebe la materia?

7.- ¿Cuál es la probabilidad de que dos o más reprueben?

8.- ¿Cuál es la probabilidad de que cuatro reprueben?

9.- ¿Cuál es la probabilidad de que uno repruebe?

10.- Obtenga la esperanza matemática de la distribución.

Problema 2. La calificación de 200 estudiantes del curso de Estadística está normalmente distribuida con media igual a 7 y desviación estándar de 0.2, en una escala de 0 a 10. Con esos datos.

1.- Construya gráficamente la curva normal correspondiente con intervalos de una desviación estándar hasta tres desviaciones estándar.

2.- ¿Cuántos estudiantes tienen entre 6.5 y 7.5 de calificación?

3.- ¿Cuál es la probabilidad de que un estudiante tenga más de 7.5 de calificación?

4.- ¿Cuál es la probabilidad de que un estudiante tenga entre 6.4 y 6.2 de calificación?

5.- ¿Cuál es la probabilidad de que un estudiante tenga entre 6.2 y 7.8 de calificación?

Problema 1 Solución.

1. Con $(q+p)^4$ tenemos

$q^4+4q^3p+6q^2p^2+4qp^3+p^4$; como. q=0.6 y p=0.4

$(0.6)^4+4(0.6)^3(0.4)+6(0.6)^2(0.4)^2+4(0.6)(0.4)^3+(0.4)^4$

0.1296+0.3456+0.3456+0.153+0.0256=1.0000

2. Interpretación de exponentes y coeficientes: hay 16 resultados posibles (ver coeficientes), donde:

q^4: hay una manera de obtener 4q’s;

$4q^3p$: hay 4 formas de obtener 3q’s y una p;

$6q^2p^2$: hay 6 formas de obtener 2q’s y 2p’s;

$4qp^3$: hay 4 formas de obtener una q y 3 p’s; y

p^4: hay una forma de obtener 4p’s.

4. Para obtener su media y su desviación estándar calculamos:

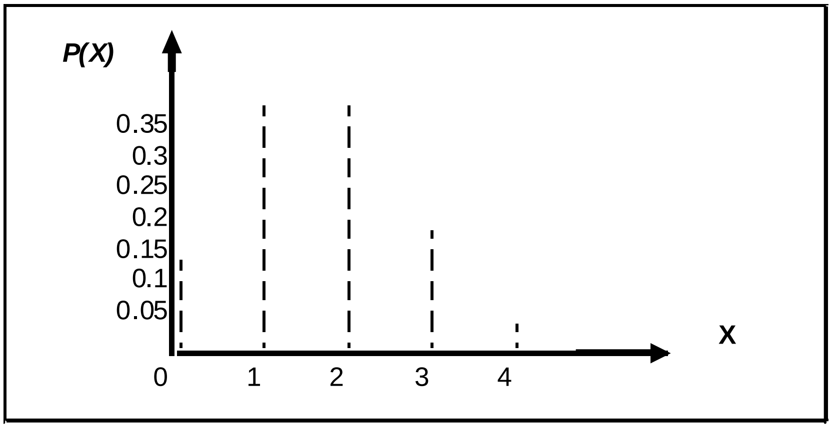
X	P(X)	XP(X)	X-μ	$(X-\mu)^2$	$(X-\mu)^2P(X)$	(X-μ)*P(X)
0	0.1296	0.0000	-1.6	2.56	0.3318	-0.2074
1	0.3456	0.3456	-0.6	0.36	0.1244	-0.2074
2	0.3456	0.6912	0.4	0.16	0.0553	0.1382
3	0.1536	0.4608	1.4	1.96	0.3011	0.2150
4	0.0256	0.1024	2.4	5.76	0.1475	0.0614
	μ=	1.5998			0.9600	0.0000

*Primera propiedad de

$$« = \sum XP(X) = 1.5998 = np = 4(0.4) = 1.6 = 1.6\sum X \text{ =E(X)}$$

$$\pm = \sqrt{npq} = \sqrt{4(0.4)(0.6)} = \sqrt{0.96} = 0.96 = \sqrt{(X-«)^2 P(X)}$$

4.Su gráfica



No es simétrica porque p=0.4 y q=0.6

Es decir p$\neq$ q

5. Transformamos X en Z: Datos discretos en continuos

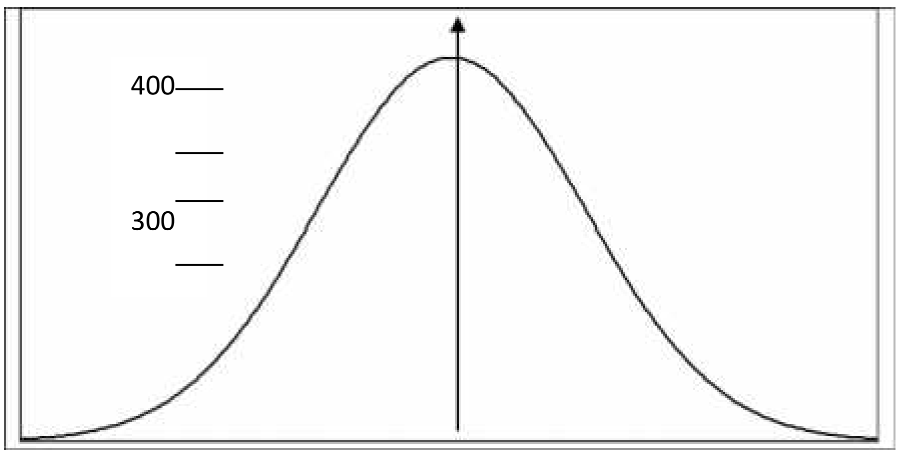
X	X-μ	$Z = \frac{X - «}{\pm}$
0	-1.6	-1.67
1	-0.6	-0.63
2	0.4	0.41
3	1.4	1.45
4	2.4	2.50

6. P(X=0)=0.1296
7. P(X$\geq$ 2)=0.3456+0.1536+0.0256=0.5246
8. P(X=4)=0.0256
9. P(X=1)0.3456
10. $E(X) = « = 1.6$

Problema 2. Solución; con N=200, μ=7 y σ=0.2

1.- Construya la curva normal

X	X-H	$Z = \frac{X - \text{«}}{\pm}$	Ordenada F(Z)	Ordenada población finita $Y_x = \frac{N}{\pm}(Z)$
7.0	0.0	0.0	0.398	398
7.2	0.2	1.0	0.241	241
7.4	0.4	2.0	0.054	54
7.6	0.6	3.0	0.0004	4



7	7.2	7.4	7.6	Valores de X originales
0	1	2	3	Valores de Z
μ	$\mu+\sigma$	$\mu+2\sigma$	$\mu+3\sigma$	Literales

2. ¿Cuántos estudiantes tienen entre 6.5 y 7.5 de calificación?

$Z_1 = \frac{6.5-7}{0.2} = \frac{-0.5}{0.2} = -2.5$; su área es 0.494

$Z_2 = \frac{7.5-7}{0.2} = \frac{0.5}{0.2} = 2.5$, su área es 0.494

P(6.5≤ 7 ≤7.5)=0.988

Luego $200(0.988) \approx$ 200estudiantes

3.- ¿Cuál es la probabilidad de que un estudiante tenga más de 7.5 de calificación?

Como Z_2=+2.5 y su área es de 0.494, la respuesta es 0.500-0.494=0.006

Ejercicio Nº. 2, para resolver.

INSTRUCCIONES: Resuelve los problemas siguientes, anotando el desarrollo de las principales operaciones y fórmulas empleadas e interpreta los resultados de cada uno de ellos según su naturaleza.

1.- En una fábrica el 50% de los trabajadores son casados, con una muestra de tres empleados, ¿cuál es la probabilidad de que:

a) Los tres son casados
b) Uno de ellos sea casado
c) Ninguno sea casado

2.- En una localidad el porcentaje de votantes por el candidato A es de 60% se toma una muestra al azar de 5 personas, ¿cuáles son las probabilidades de que en dicha muestra, voten por el candidato mencionado?

a) Ninguna persona
b) Más de 3 personas
c) Cuando menos 3 personas

3.- El 3% de los tornillos que produce una máquina son defectuosos, ¿cuál es la probabilidad que de 100 tornillos escogidos al azar cuando mucho haya dos defectuosos?

4.- Se ha comprobado que el 2% de una caja que contiene 200 pilas, son defectuosas ¿cuál es la probabilidad que exactamente 3 de ellas sean defectuosas?

5.- La media de los diámetros interiores de una muestra de 200 rondanas, producidas por una máquina es de 0.502 pulgadas y su desviación estándar de 0.008 pulgadas, el propósito para que se destinan estas rondanas permite una tolerancia máxima en el diámetro de 0.496 a 0.508 pulgadas.

De otra manera las rondanas se consideran defectuosas.

a) Si los diámetros se distribuyen normalmente construye la gráfica representativa con intervalos de 1/2 de desviación estándar hasta tres desviaciones estándar.
b) Determinar el tanto por ciento de rondanas defectuosas producidas por la máquina.
c) ¿Cuál es la probabilidad de que al seleccionar una arandela, su diámetro sea mayor que 0.510 pulgadas?

6.- El tiempo de duración de 5,000 pilas secas para focos fotográficos producidos por una compañía está normalmente distribuidos con media igual a 800 minutos y desviación estándar igual a 40 minutos.

a) Construya gráficamente la curva normal correspondiente con intervalos de 1/2 de desviación estándar hasta tres desviaciones estándar.

b) ¿Cuántas pilas duran entre 780 y 820 minutos?

c) ¿Cuál es la probabilidad de que al seleccionar una pila esta dure cuando menos 750 minutos?

GENERACION DE LA DISTRIBUCION NORMAL CON EXCEL (MARTINEZ, 2005: 95)

- **DISTRIBUCIÓN NORMAL**

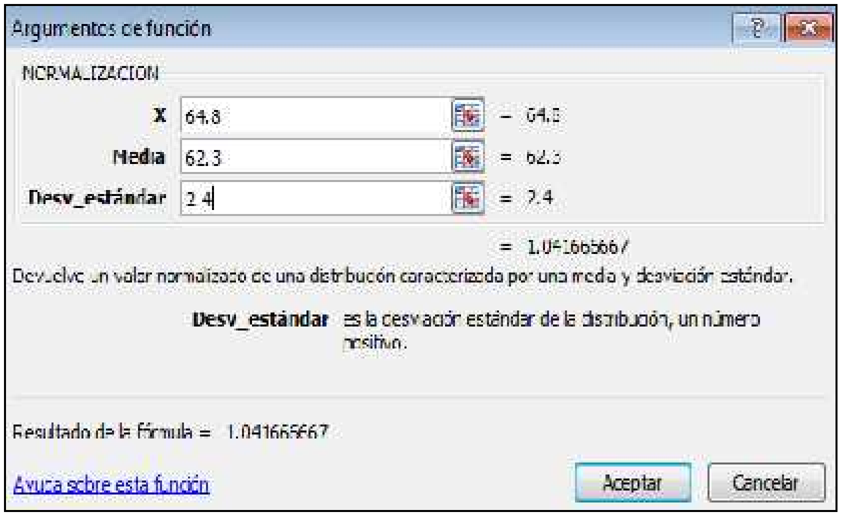
Fórmula: $z = \dfrac{x - \mu}{\sigma}$

A. Cuando solamente se requiera calcular el valor correspondiente a Z, se procederá de la siguiente forma:

- Supongamos que se tiene la información:

$x = 64{,}8 \qquad \mu = 62{,}3 \qquad \sigma = 2{,}4$

- Colocamos el cursor en **fx/** click **/** insertar una función/ *estadísticas* / clic/ **NORMALIZACIÓN**/ click/ aceptar/argumentos de función: x=64.8; media=62.3; en y en *desv_estándar=2.4*./ aceptar.



X

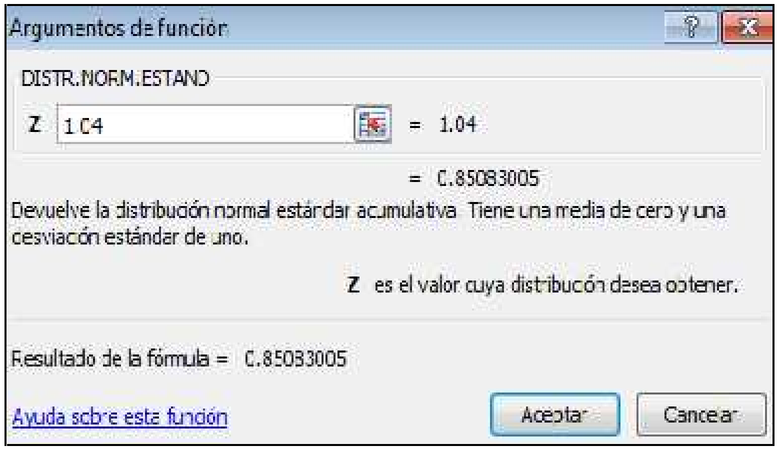
62,3 64,8

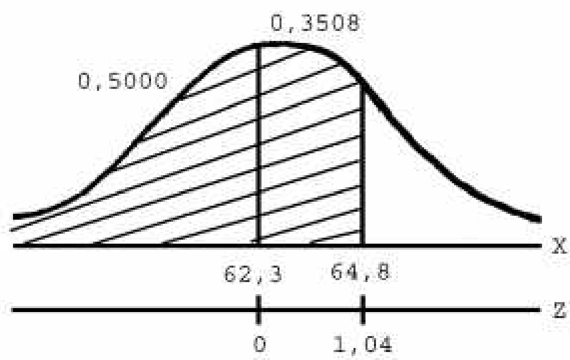
Z

0 1,04

B. Pero, si el objetivo de nuestro cálculo es obtener el ÁREA BAJO LA CURVA NORMAL, utilizamos la función correspondiente, en este caso será DIST.NORM.ESTAND.

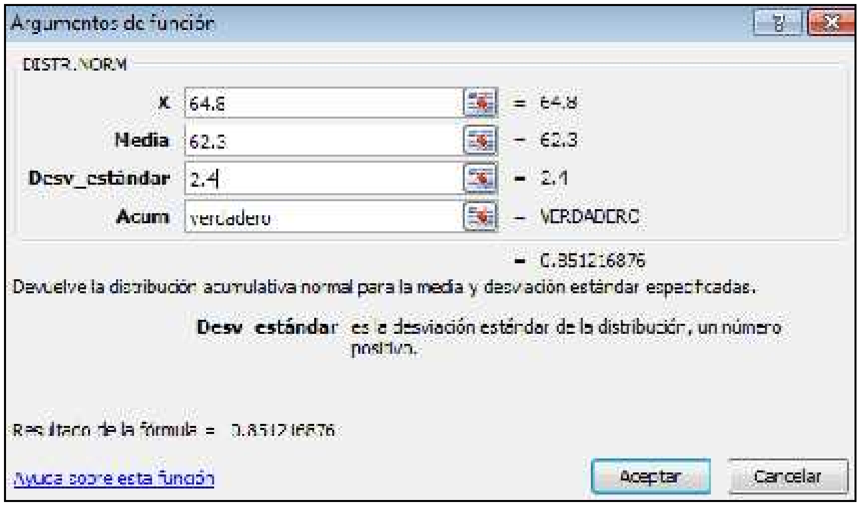
- Colocamos el cursor en **fx/** click **/** insertar una función/ *estadísticas* / clic/ **DIST.NORM.ESTAND**/ click/ aceptar/argumentos de función: z=1.04/ aceptar





C. La forma DIST.NORM, nos permite agilizar las operaciones realizadas cuando utilizamos el procedimiento (B), para hallar el área bajo la curva normal, así:

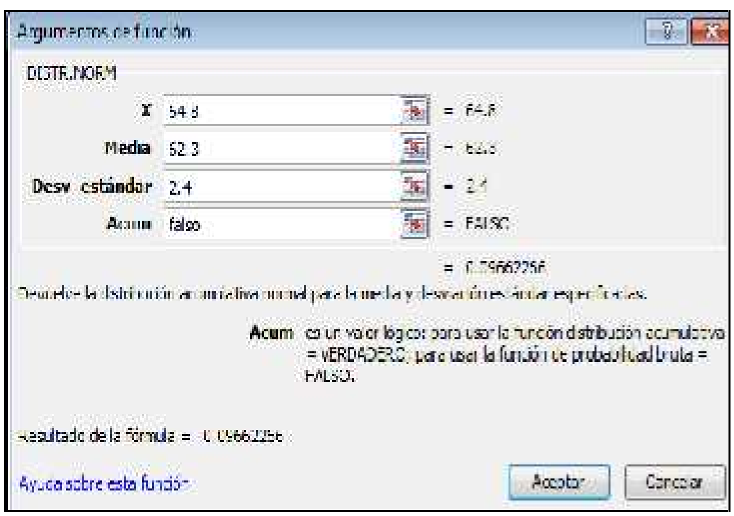
- Colocamos el cursor en **fx/** click **/** insertar una función/ *estadísticas /* clic/ **DIST.NORM.**/ click/ aceptar/argumentos de función: x=64.8; media= 62.3; desv_estándar=2.4; acum= verdadero/ click/ aceptar



Esta es la probabilidad correspondiente a **la función distributiva acumulativa:**

$P_{(x < 64,8)}$ = 85,12%

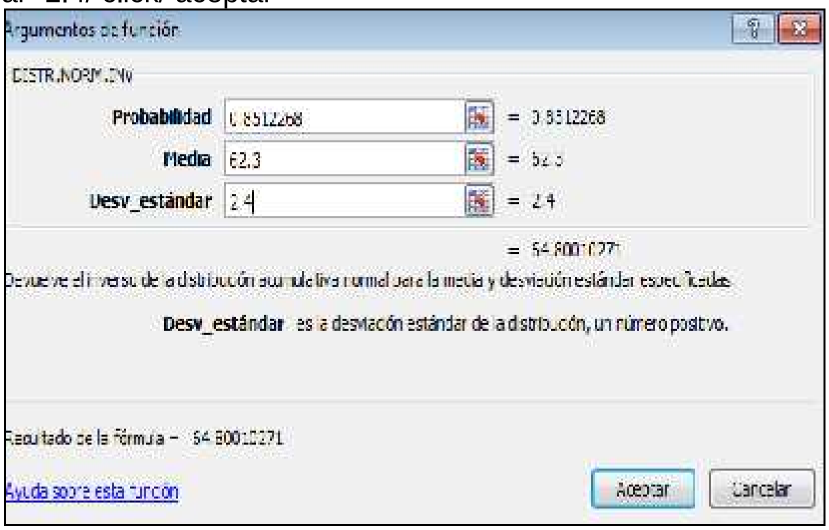
- Si en el valor lógico ACUMULATIVO, digitamos la palabra FALSO nos dará como resultado 0,09662266 (9.66%), correspondiente a la **función de probabilidad bruta.**



D. Ahora, otro procedimiento que podemos realizar, conociendo la probabilidad correspondiente al área bajo la curva normal, además de la MEDIA y DESVIACIÓN TIPICA, podemos determinar el valor x, utilizando para ello la función DIST.NORM.INV.

Veamos:

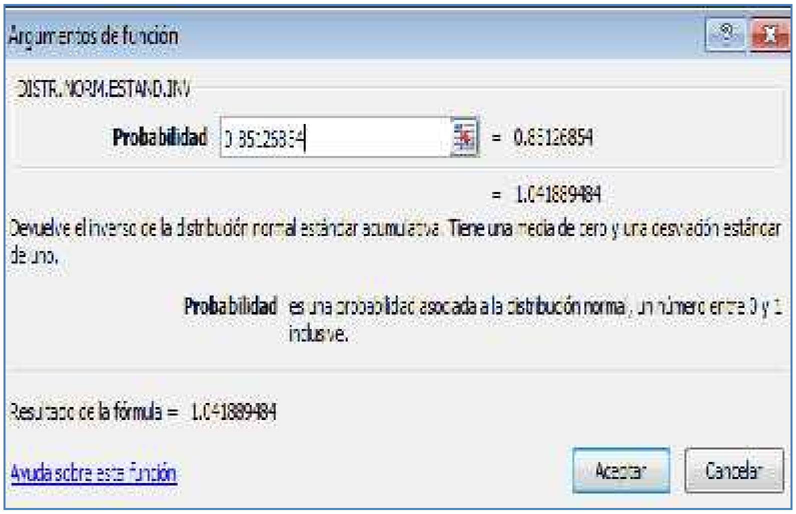
- Colocamos el cursor en **fx/** click **/** insertar una función/ *estadísticas* / clic/ **DIST.NORM.INV**/ click/ aceptar/argumentos de función: prob=0.8512268; media= 62.3; desv_estándar=2.4/ click/ aceptar



- Observamos que el valor de x = 64,8, tal como lo habíamos considerado con el ejercicio D

E. Algo similar lo podemos realizar con la FUNCIÓN correspondiente a la DISTR.NORM.STAND.INV para determinar el valor de Z, digitando el valor correspondiente de la probabilidad conocida.

- Colocamos el cursor en **fx/** click **/** insertar una función/ *estadísticas* / clic/ **DIST.NORM.STAND.INV**/ click/ aceptar/argumentos de función: prob=0.8512268/ click/ aceptar



Observemos, que el resultado obtenido en este procedimiento es de 1,041889846, para el valor de z; (z = 1,04)

APÉNDICE C: INFERENCIA ESTADÍSTICA.

Conceptos básicos:

- **Muestreo**

Importancia: Una vez que definimos, explicamos e ilustramos el concepto de probabilidad, la cual vimos que constituía el eje rector para hacer análisis económico ampliado, sobre la base de la estadística descriptiva que sienta las bases para introducirse y profundizar usando los métodos de la **inferencia estadística**, la cual se basa en el análisis de una muestra para inferir las características de la población de la que proviene.
Lo anterior fue muy valioso porque a partir de la *naturaleza* y número de *resultados posibles* que se generan en un experimento, con una probabilidad de ocurrencia dada de manera ex-ante, pudimos constituir el marco muestral de los mismos, con el que pudimos calcular la probabilidad de ocurrencia de cada uno de ellos, al igual que la probabilidad de ocurrencia de un número determinado de ellos, que fueran de interés particular para el investigador. También al contar con el marco muestral pudimos decidir cómo agruparlos, sobre todo cuando son muchos y ya no es fácilmente visible el marco muestral (con las fórmulas de las permutaciones y de las combinaciones). Este agrupamiento constituye la base para que el investigador diseñe el método de muestreo (selección de la muestra) que le permita cumplir con los objetivos de su estudio. En este contexto es que también al saber cómo podemos relacionarlos o agruparlos es que a partir del análisis combinatorio, con sus fórmulas, pudimos crear distribuciones probabilísticas discretas y continuas, cuya tipificación estadística fundamentó el rigor técnico con que se puede usar la inferencia estadística, así como para visualizar en que variables económicas son susceptibles de aplicar cada una de ellas (binomial, poisson, normal, etc.) El entendimiento de los conceptos y aplicación de los métodos de la estadística inferencial ampliaron la cobertura de esta disciplina en apoyo del sólido análisis que debe hacerse de las variables que integran la economía nacional o de la empresa.

Continuando con su aplicación, ahora veremos cómo se usa para la obtención de muestras probabilísticas, que obtendremos de poblaciones finitas e infinitas. Motivo por el cual es conveniente introducir de manera formal la definición de los siguientes conceptos:

- **CONCEPTO DE UNIVERSO Y MUESTRA:**

UNIVERSO O POBLACIÓN ESTADISTICA: Se define como el conjunto de elementos que poseen la característica que el investigador desea estudiar o simplemente como la suma de las unidades elementales.

Si el número de unidades elementales es igual al número de observaciones; se dice que la población es la suma de las observaciones.

Por ejemplo: Si hay 600 personas e interesa la variable “x”, el peso en kilogramos de las personas, cada persona es una unidad elemental y por lo tanto la población son las 600 personas.

El **tamaño de una población** se representa generalmente por **N**. Luego, una población en sentido estadístico es un conjunto de elementos (generalmente definida) que puede conocerse por medio de un análisis completo y exhaustivo.

La población puede ser: **FINITA** o **INFINITA**.
El ejemplo de las 600 personas ilustra una población FINITA.
Una población INFINITA podrá referirse por ejemplo al número de moscos que hay en el mundo entero.

Cada una de las unidades elementales, tiene varias características identificables y numerables; es decir que cada característica puede representarse por un número.
Por ejemplo: Si la población es de animales, sus características pueden ser:
Su peso; la dieta a que están sujetos; su producción (según su clase: vacas, gallinas, etc.)

En la teoría de la probabilidad moderna, una población se representa gráficamente en la siguiente forma:

```
*    **    *      *     *
        *      **     *          *    *
          *    *      *     *        *    *
```

(R, Q, P)

Donde R representa el conjunto-Universo o población;
Q es una álgebra de Boole;
P es la medida de la probabilidad dentro de la población.
Una muestra es un conjunto de n observaciones-unidades elementales-extraídas de la población. Esta n es el tamaño de la muestra.
Si en el ejemplo anterior se seleccionan 20 personas de la población de 600, se ha tomado una muestra de tamaño: n = 20.

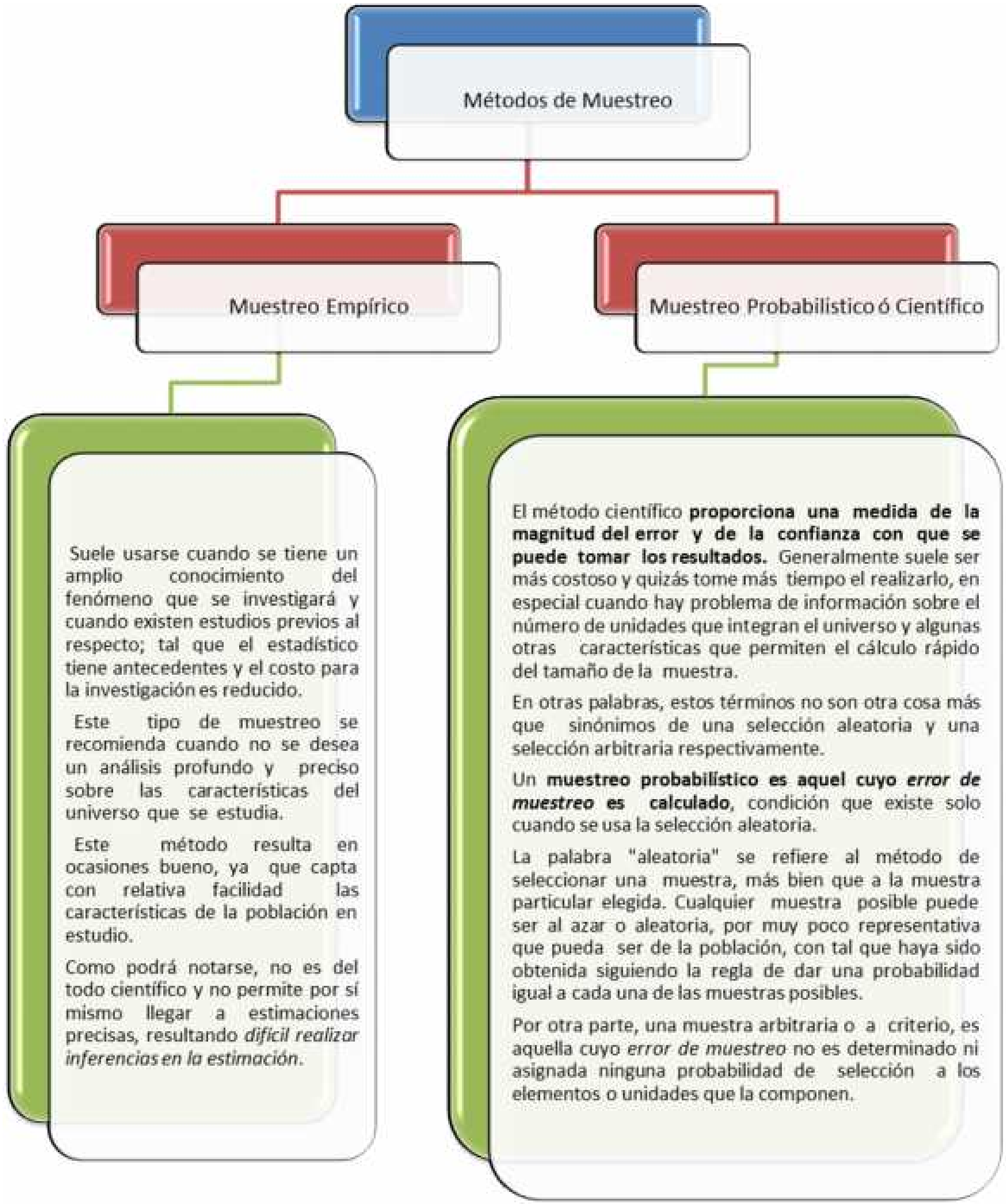
MUESTRA*: se define como una porción de la población estadística que es seleccionada para estudiarla y de ella inducir (inferir) las características que puede tener la población estadística.*

El tipo de muestra y representatividad que se obtiene con ella depende de la forma en que haya sido extraída la muestra de la población. Se habla de procedimientos “dirigidos” (también conocidos como piloto) y de métodos de selección probabilística como el muestreo simple aleatorio, de muestreo sistemático, de muestreo estratificado, por conglomerados, etcétera.

Dentro de los primeros se habla de una **muestra no aleatoria**, que es la *parte de la población que el investigador selecciona arbitrariamente para estudiar a partir de ella a la población.*

Cuando se utilizan los segundos métodos se habla de una **muestra aleatoria** y se dice que tienen en común el hecho de que *seleccionan la muestra al azar, además de que cada elemento de la población tiene la misma probabilidad de ser incluido en la muestra, que se conoce como muestra probabilística y tiene características importantes* que más adelante describiremos.

- **METODOLOGIA DEL MUESTREO ESTADÍSTICO.** [11]



Se recomienda el uso del **muestreo científico**, también llamado aleatorio o probabilístico, es que por lo regular el investigador tiene a su disposición muchas muestras con composiciones diferentes, de manera que según la muestra utilizada puede haber errores o diferencias entre los valores de los parámetros y de sus estimadores correspondientes. Dichos errores no se pueden evitar en una selección aleatoria pero si cuantificar a priori cuando se determina matemáticamente el tamaño de la muestra, es decir, en la planeación de la investigación por muestreo se conoce el *error de muestreo* (que debe ser igual o menor al error permitido), así como el nivel de confiabilidad de los “estadísticos” muestrales.

ERRORES DE MUESTREO Y DE NO MUESTREO.

La exactitud o confiabilidad de una muestra, depende de dos tipos básicos de errores: **errores de muestreo**, *que se reflejan en estimaciones matemáticas de la precisión de estimadores provenientes de muestras particulares*, y *se manifiestan en diferentes* formas clasificadas bajo la notación de *sesgos o distorsiones*.
Los errores de muestreo *se miden a través de las llamadas fórmulas de error estándar*. De acuerdo con estas fórmulas, se hacen estimaciones de la precisión de estimadores muéstrales particulares y siguiendo el procedimiento apropiado; estas mismas fórmulas sirven de base para determinar el tamaño de la muestra requerida, de acuerdo con una precisión especificada previamente. Las fórmulas del error estándar han sido desarrolladas para una gran variedad de diseños muéstrales y en la actualidad es una cuestión rutinaria su aplicación a cada uno de los casos.

Los errores de muestreo surgen de la variación en los estimadores provenientes de distintas muestras del mismo tamaño.
El valor de los errores determina la precisión con que los valores muéstrales estiman a los parámetros poblacionales.
La probabilidad de que un parámetro esté contenido dentro de un cierto rango construido alrededor de los diferentes estimadores muestrales, se obtiene por medio de la teoría de la probabilidad para distintos diseños muéstrales.
Así, con base en esta teoría, el margen de error (o error de muestreo) que se puede esperar con un diseño de muestreo y tamaño de muestra determinados, se puede calcular a diferentes niveles de precisión bajo el supuesto de una selección aleatoria, la cual requiere que cada miembro de la población tenga la misma probabilidad de ser seleccionado. Una vez que se conocen el error estándar y la precisión buscada, se pueden calcular: el tamaño de la muestra y los recursos necesarios para la investigación.
Contrariamente, el tema de los errores de no muestreo es a la fecha un tema que requiere una vasta experiencia y la cual es ajena a la disciplina matemática.

Incluidas en el concepto de errores no de muestreo, están las innumerables influencias que tienden a distorsionar o sesgar los estimadores provenientes de la muestra, la selección arbitraria de los miembros de la muestra, fraseo perjudicial en las preguntas, actitudes preconcebidas por el entrevistador y muchos otros factores pueden producir valores muestrales que no representaran a los valores de los parámetros de la población, no importa que tan grande sea la muestra.
Distintos a los errores de muestreo, éste tipo de sesgo es independiente del tamaño de la muestra.

SELECCIÓN DE LA UNIDAD DE MUESTREO.

La aplicación de los métodos de muestreo estadístico tiene por objeto, seleccionar algunos elementos del universo que se trata de estudiar, para poder hacer inferencias sobre sus características. La selección de estas unidades se hace a partir de una lista, mapas, croquis, directorios, o una combinación de estos elementos informativos, los que deben contener todas las unidades de interés y permitir determinar la probabilidad de su inclusión; así mismo, que en el momento de levantar la encuesta, la identificación de cada unidad en la muestra sea hecha sin ninguna ambigüedad.

Al conjunto de todos los elementos se le llama: **MARCO MUESTRAL**.

De acuerdo a la forma de seleccionar estas unidades se pueden dar las siguientes maneras de hacerla:

Con Reemplazo:

Se aplica cuando una unidad de muestreo previamente seleccionada para ser incluida en la muestra, se regresa a la población para concursar y poder ser nuevamente seleccionada para volver a ser parte de la muestra definitiva.

Sin Reemplazo:

En este caso la unidad de muestreo, una vez extraída para ser parte de la muestra, ya no es regresada a la población para concursar y ser seleccionada nuevamente para formar parte de la muestra definitiva.

Etapas:

A las unidades contenidas en el universo muestral posiblemente convenga agruparlas, a estos grupos a su vez, se volverlos a agrupar y así sucesivamente. Dependiendo del número de agrupamientos de las unidades de interés (o últimas unidades de muestreo), es el nombre que se les da. Por ello, si el marco muestral no presentó agrupamientos, el muestreo se llamará monoetápico, que es la selección directa de las unidades de interés. Si el marco muestral presenta agrupamientos de un sólo orden se llamará bietápico, o lo que es lo mismo se seleccionarán primero los grupos de unidades (de primera etapa) y finalmente se seleccionarán los de interés o de segunda etapa, y así sucesivamente se tendrá el muestreo trietápico, tetraetápico, etc.

Probabilidad:

Si las unidades de muestreo en cada etapa son seleccionadas con la misma probabilidad, el muestreo se llamará **equiprobable**; en el caso contrario se dice que es de probabilidades variables de selección en la ó las etapas que correspondan.

Estratos:

La precisión al hacerse las estimaciones básicamente depende de dos factores:

a) Del tamaño de la muestra; y
b) De la variabilidad o heterogeneidad de la población.

Es evidente que mientras más grande sea la muestra, representará más fielmente a la población, tal que se pueden mejorar las estimaciones aumentando el tamaño de la muestra. En cuanto al segundo factor para aumentar la precisión, puede dividirse el marco muestral, (si es que se dispone de los medios necesarios) *en* **clases homogéneas** *llamados* **estratos** *y seleccionar separadamente en cada estrato una muestra*, garantizando con esta forma cualquier representación deseada de todos los estratos de la población. La denominación de un método de muestreo se forma indicando estos conceptos: etapa, probabilidad y con ó sin reemplazo. Al constituir los estratos con elementos cuyos valores son cercanos entre sí, se garantiza que la variabilidad entre ellos sea reducida y por consiguiente, que el error de muestreo que se obtiene con este método sea el menor con respecto a otros métodos de muestreo utilizados; esto se demuestra en las siguientes secciones.

MANEJO DE LAS TABLAS DE NÚMEROS ALEATORIOS

La selección de las unidades de muestreo debe hacerse basándose en las leyes del azar; esto es, debe asignarse a cada unidad del marco muestral una probabilidad de inclusión en la muestra. Con este método la muestra se obtiene en selecciones sucesivas de una unidad, cada una con una probabilidad asignada de antemano, según sea el modelo de muestreo que se utilice, hasta completar el número de unidades que deben incluirse en la muestra para cada etapa. Un procedimiento práctico para seleccionar las unidades, es utilizando una tabla de números aleatorios como la que aparece en el apéndice N de la sección de tablas estadísticas.

CONSTRUCCIÓN DE LAS TABLAS DE NÚMEROS ALEATORIOS

Conviene destacar que estas tablas sirven para asegurar que todos los elementos del universo tienen la misma probabilidad de ser seleccionados aleatoriamente como parte de la muestra que se extraiga de la población.

Estas tablas están constituidas por arreglos de dígitos ordenados de manera tal que cada uno de ellos represente a un elemento de la población e indica que dicho elemento tiene la misma probabilidad que el resto de ser seleccionado aleatoriamente como integrantes de la muestra.

Estas tablas se construyen de diferentes maneras:
- Usando la computadora de manera similar al proceso de la ruleta.
- Usando ciertas funciones matemáticas; ó
- Usando instrumentos mecánicos basados esencialmente en el principio de la ruleta.

El uso de las tablas de números aleatorios puede ilustrarse con el siguiente ejemplo, relativo a la **selección aleatoria** de la muestra.

Supóngase que se van a seleccionar tres Escuelas de Medicina Veterinaria y Zootecnia para ser consideradas como muestra de las 18 escuelas de Medicina Veterinaria y Zootecnia existentes en el país:

Si n = 3 y N = 18. Decimos que el universo está constituido por dos dígitos; si N fuera 4327, diríamos que está formado por cuatro dígitos; el número de dígitos del universo es el límite máximo para trabajar dichas tablas. Así, en nuestro ejemplo, se hace la relación o numeración de las escuelas que integran el universo: a cada uno de las 18 Escuelas se le asigna un número de dos dígitos: 01, 02, 03, ..., 18.
En seguida se seleccionan pares de números de la tabla de manera consistente.
Por ejemplo: la selección podría empezar en la parte superior de la tabla, - primera columna -, la siguiente columna, etc. Esto produce los siguientes pares de dígitos: 01, 04, 06.

Estos dígitos identifican la escuela en la población que será considerada como elemento de la muestra.

Si el número par al azar excede el número de unidades posibles de muestreo (N = 18) como el número 31, el número es ignorado y se selecciona el siguiente número, 16 -por ejemplo- y al seguir seleccionando para completar el tamaño de la muestra y ésta vuelve a aparecer, en este caso también se ignora y se continúa buscando un número distinto a 16 y no mayor que 18.

De esta manera se obtienen las tres escuelas que formarán la muestra. Esta no es la única manera para seleccionar pares de dígitos en la tabla de manera horizontal, diagonal, en zig-zag, etc. Lo importante es que el procedimiento sea consistente.

El segundo medio de selección probabilística, el **sistemático,** es en esencia una simple variante del procedimiento anterior. Implica la selección de las unidades de la muestra de manera sistemática empezando con uno de los dígitos, dicho en otras palabras, *la selección de cada uno de los elementos de la muestra es a intervalos regulares, una vez que fue escogida la primera de n unidades que constituirán la muestra.*
Esto es, si hay N unidades muestrales en la población, y se desean n para la muestra, cada N/n unidad es seleccionada, empezando con un número aleatorio.
Así usando el ejemplo anterior cada sexta unidad será seleccionada: (N/n = 18 / 3 = 6) empezando con un número aleatorio entre 01 y 06 inclusive. Este número aleatorio se puede obtener también de la tabla de números aleatorios.

- **METODOS DE MUESTREO**

Los métodos de muestreo tienen por objeto indicar la forma como se seleccionará el número de unidades que deben incluirse en la muestra. Dependiendo de la forma en que estas se seleccionan, de la confianza estadísticamente hablando que se requiera al hacer las inferencias de los resultados muestrales como estimadores de los valores poblacionales y del error de muestreo que se determine en relación al error permitido, es como se denomina el método de muestreo a utilizar en la realización del estudio.

MUESTREO SIMPLE ALEATORIO

Recordando que por muestreo probabilista se entiende un método de muestreo en el que cada miembro de la población tiene una probabilidad conocida de ser incluida en la muestra, decimos que cuando todos los miembros de la población tienen la misma probabilidad de ser seleccionados se denomina muestreo simple aleatorio.

Ejemplo: Si una caja contiene seis pedacitos de papel numerados del 1 al 6; si se desea elegir una muestra de la caja de tamaño 3, sin reemplazo, el muestreo simple aleatorio indica que la probabilidad de cada uno de los 6 papelitos es 1/6. Al extraer el segundo, la probabilidad de cada uno es 1/5 y así sucesivamente. En este caso cada número dentro de la caja tiene la misma probabilidad de ser seleccionado.
Esto es, la probabilidad del número 4 es igual a:
P (A o B o C) = 1/6 + 5/6 x 1/5 + 5/5 x 4/5 x 1/4 = 3/6.

En general, se puede decir que si el tamaño de la muestra es **n** y el de la población **N**, en el muestreo simple aleatorio, cada miembro de la población tiene una probabilidad de encontrarse en la muestra de **n/N** .
Por ejemplo: Si de entre 120 estudiantes se seleccionan 10 al azar y todos tienen la misma probabilidad de ser elegidos, cada uno de los 120 estudiantes, tiene una probabilidad de 10/120 de estar en la muestra.
Ahora ¿cuál es la probabilidad de seleccionar una muestra de tamaño n a partir de una población de tamaño N?
Suponiendo de N = 6 y n = 3:

$$\begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{n} \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \frac{6!}{3!(6-3)!} = \frac{6!}{3!*3!} = 20 \text{ muestras posibles}$$

Respuesta: Cuando se adopta el muestreo aleatorio simple cada muestra tiene igual probabilidad de ser seleccionada y es de 1/20.

En general, se dice que cuando se selecciona una muestra de tamaño n, a partir de una población de tamaño N por muestreo simple aleatorio la probabilidad de que se seleccione una cualquiera de las $\begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{n} \end{bmatrix}$ muestras posibles será: $\dfrac{1}{\begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{n} \end{bmatrix}}$

Lo anterior se refiere a los casos en que el muestreo se realizo sin reemplazo. Lo mismo sucede cuando se realiza con reemplazo, aunque en la práctica se utiliza generalmente el muestreo sin reemplazo.

MUESTREO ESTRATIFICADO [11]

De acuerdo con este método, la población se divide en estratos basados en características consideradas relevantes para el sujeto bajo estudio, y se seleccionan las unidades de muestreo de cada uno de los estratos.
Por ejemplo: investigando tiendas al menudeo en la ciudad de Cuernavaca; las tiendas en la ciudad podrán clasificarse primero por tipo de tienda (abarrotes, farmacias, etc.) y luego por tamaño de tienda. Para cada estrato, tipo o tamaño de tienda, se puede estimar el número de tiendas y calcularse cuántas de estas tiendas -unidades de muestreo- deben incluirse en la muestra. Es común en tales casos, seleccionar la mayoría de las unidades de muestreo de los estratos conteniendo las tiendas grandes y sólo una pequeña proporción de unidades de muestreo de los estratos que contienen relativamente pocas tiendas.
Para que sea útil el muestreo estratificado se deben reunir las siguientes tres condiciones:
1) Deben conocerse ciertas características relevantes que influencian fuertemente el fenómeno bajo estudio:
2) Que la población sea susceptible de dividirse de acuerdo con las características relevantes:
3) La división relativa de la población debe conocerse con cierto grado de precisión.
Una muestra estratificada puede obtenerse aún cuando no se pudieran identificar los elementos del estrato, siempre y cuando se conozca después de haberse seleccionado la muestra. El problema sin embargo, es que los errores de muestreo de las estimaciones resultan mayores que si se hubiera estratificado antes.

Si el número de unidades de muestreo seleccionadas de cada estrato es proporcional al tamaño relativo del estrato en la población, el resultado es una muestra estratificada proporcional, lo contrario es una muestra estratificada no proporcional. Esto último es preferible si los diversos estratos no son homogéneos con respecto a la característica bajo estudio.

El error de muestreo de una muestra estratificada puede considerase menor que el de una muestra simple aleatoria del mismo tamaño. Lo anterior se debe a que el diseño de estratificaciones hace uso de información adicional, considerando la división de la población de acuerdo con las características relevantes y sirve para reducir el margen de error de muestreo.

El problema con este método, es que aún cuando se conocen las características relevantes y en base a ellas se estratifica, el tamaño relativo de los estratos en la población no siempre se conoce con gran exactitud.

Debido a esta escasez de información, las ventajas obtenidas con la estratificación se pierden con las variaciones introducidas por la información incorrecta referente al tamaño de los estratos en la población, elemento que desafortunadamente se subestima frecuentemente.

Los diseños de estratificación se pueden combinar con otras como por ejemplo:
- Muestreo por área; y
- Los esquemas de muestreo por conglomerados o racimos.

Ejemplo de la situación anterior podría ser el siguiente: digamos que México podría subdividirse en estratos regionales, tales como:

- Norte;
- Sur;
- Este: y
- Oeste.

Con áreas seleccionadas dentro de cada uno de estos estratos o regiones y con miembros de la muestra seleccionados al interior de cada una de estas áreas, en grupos o "racimos". Similarmente, la selección de los miembros de una muestra estratificada podría realizarse, ya sea usando procedimientos aleatorios o arbitrarios.

MUESTREO POLIETAPICO

Este método requiere la selección de las unidades de muestreo en diferentes etapas, existiendo unidades de primera, segunda, etc.., etapa en un diseño muestral.

Por ejemplo: si el interés es conocer la opinión de los médicos veterinarios zootecnistas sobre los programas de estudio de las diferentes escuelas y facultades de Medicina Veterinaria y Zootecnia y si para ello se decide realizar la investigación en la ciudad de México, entonces la clasificación de la ciudad en distritos permite obtener la unidad de primera etapa; la clasificación en colonias es la unidad de la segunda etapa; la selección de las manzanas a muestrear es la unidad de tercera etapa; y la selección aleatoria de los médicos residentes en las manzanas previamente seleccionadas, constituyen la unidad de cuarta etapa.

MUESTREO POR ÁREAS

Cuando la población se distribuye sobre un área muy grande, la selección de los elementos de la muestra de toda el área puede resultar un procedimiento ineficiente y costoso. Esto es particularmente cierto, si a las personas que entrevistan se les paga por hora y la mayor parte del tiempo se va en viajar. El muestreo por áreas fue diseñado para resolver este problema. Se basa en una subdivisión apriori de la población en áreas; la selección de algunas de estas áreas con la ayuda de los métodos de muestreo aleatorio y la restricción a la selección de las unidades que integrarán la muestra, solamente en esas áreas.

La restricción geográfica sirve para concentrar los esfuerzos de trabajo en ciertas regiones, provocando reducciones sustanciales en el costo del trabajo de campo en comparación a una muestra del mismo tamaño proveniente de un diseño distinto al de áreas.
Esta técnica de muestreo puede usarse para trabajar con muestras irrestrictas y estratificadas. De hecho en investigaciones de gran escala la técnica de estratificar áreas es generalmente la regla, porque asegura la representatividad de todos los segmentos relevantes de la población a costos bajos.
En cada investigación el diseño de áreas se realiza en varias etapas; cada etapa sirve para restringir el área geográfica de la cual se seleccionarán las unidades de la muestra.

- **Muestreo por conglomerados**.

Con este método se distribuyen los elementos de la población estadística en varios grupos o conglomerados, de manera que cada uno de estos grupos tenga un número de cuentas con diferentes valores; ello con objeto de que cada grupo sea una réplica del universo o la población estadística. Una vez realizado el agrupamiento de datos, se procede a enumerar los grupos o conglomerados constituidos y de ellos se selecciona uno al azar para que constituya la muestra que servirá para realizar la investigación deseada.

3.- APLICACIONES

Sabiendo que una ***encuesta*** *es una investigación que realiza el experto para obtener datos de interés específico sobre un tema determinado, a continuación se expone cómo se realiza dicha investigación ilustrando la manera cómo obtener los datos utilizando algunos métodos de muestreo, como los siguientes:*

- **APLICACIÓN DEL MUESTREO SIMPLE ALEATORIO**

Ejemplo 1:

Aún cuando este método es el más simple de los clasificados como probabilísticos, su sencillez no deja de ser útil para ilustrar las ventajas que se derivan de la aplicación de esta metodología al análisis de fenómenos económicos, al igual que los demás métodos de muestreo estadístico, se caracteriza por proporcionar estimación de los caracteres de la población.
Se asigna igual probabilidad de selección a cada unidad perteneciente a la población. Si N es el número de unidades, la probabilidad de selección de cualesquiera de ellas es: 1/N.

En un muestreo sin reemplazo el número de muestras distintas de tamaño n, sacadas de las N unidades de la población está dado por:

$$\begin{bmatrix} N \\ n \end{bmatrix} = \frac{N!}{n!(N-n)!}$$

Los estimadores obtenidos con la muestra que se use serán insesgados cuando su esperanza matemática sea igual al parámetro poblacional:

$$E(\bar{y}) = \bar{Y}$$

Demostración: $\bar{y} = \frac{1}{n}\sum y_i$

$E(\bar{y}) = \frac{\sum (y_1 + y_2 + \ldots + y_n)}{n} = \frac{E(y_1) + E(y_2) + \ldots + E(y_n)}{n} = E(\bar{y}) = \frac{n\bar{Y}}{n}$ por lo tanto $E(\bar{y}) = \bar{Y}$

El estimador del total de la población definido por $\hat{Y} = N\bar{y}$ es insesgado porque:

$$E(\hat{Y}) = E(N\bar{y}) = NE(\bar{y}) = NY = Y$$

Aplicaciones: Para ello se supone que se conoce el tamaño de la muestra requerida, el cual se estudiará posteriormente en detalle.

Objetivo: Se desea estimar el total de familias en la localidad "y" con una muestra simple aleatoria cuyo tamaño está dado por cuatro manzanas.

Notación:

F = n/N= Fracción de muestreo.

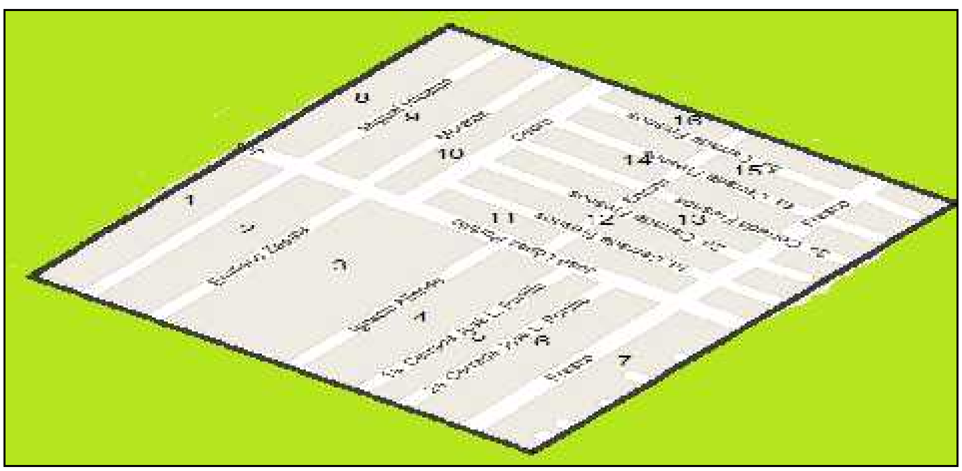
N = Número de manzanas en la localidad.

$\hat{Y}$ = Población total estimada

$\bar{y}$ = Promedio de familias por manzana en la muestra

m = Promedio de personas por familia.

El mapa de la localidad revela la siguiente distribución de las manzanas.



Las manzanas se numeran siguiendo un orden determinado: ascendente o descendente en este caso, resultaron ser 16 en total.

Conociendo N = 16 y n = 4 se seleccionará la muestra con la tabla de "números aleatorios". Suponiendo que las manzanas seleccionadas son:
los número 16, 3, 9 y 11.
En seguida, se hace un listado de las manzanas seleccionadas registrando el número de familias que existen en cada una de ellas. Los resultados son:

La manzana 16 tiene 4 familias.
La manzana 3 tiene 9 familias.
La manzana 9 tiene 9 familias.
La manzana 11 tiene 10 familias.

Recordando que el total de familias se estima por:

$\hat{Y} = N\bar{y};$ si N=16 y $\bar{y} = \frac{1}{n}\sum y_i = \frac{1}{4}(4+9+9+10) = \frac{32}{4} = 8$

Tendremos que $\hat{Y}$ =16(8); $\hat{Y}$ =128 familias en la localidad.

Se puede estimar que el cálculo del total de las familias en la localidad tenga un 95% de probabilidad de haber caído en el intervalo de confianza con la siguiente fórmula:

$$N\bar{y} - \frac{tNs}{\sqrt{n}} * \sqrt{1-F} \leq \hat{Y} \leq N\bar{y} + \frac{tNs}{\sqrt{n}} * \sqrt{1-F}$$

Donde t es el valor de la normal desviada correspondiente a la confianza de probabilidad deseada cuando n es menor que 30 y **s** es la varianza muestral.

Como se recordará:

Con ~ = 5% y un número infinito de grados de libertad se halla en tablas $t_{\sim}$ = 1.96; sabemos que:

$$S^2 = \frac{\sum(y_i - \bar{y})^2}{n} = \frac{\sum y_i^2}{n} - (\bar{y})^2 = \frac{278}{4} - 8 = 5.5$$

Como $S = \sqrt{s^2} = \sqrt{5.5} = 2.3$ y $t_{\sim}$ = 1.96 tendremos que

Limites de confianza $= \frac{16(32)}{4} \pm \frac{(1.96)(16)(2.3)}{\sqrt{4}} * \sqrt{1 - \frac{4}{16}} =$ 125 a 131

El total estimado de familias (128) se halla entre 125 y 131 con una seguridad o confianza del 95%.

El número total de habitantes se puede saber multiplicando el total estimado $(\hat{Y})$ por el promedio de personas por familia (m)

Si m= 5.4; $\hat{Y}$ = 128

m $\hat{Y}$ = 5.4 (128) = 691 habitantes en la localidad "gama"

Ejemplo 2:
Si deseamos hacer un estudio de mercado sobre el interés que puedan tener los consumidores sobre un producto o servicio determinado, hacemos lo siguiente:

Paso 1.- Se recaban datos para calcular la muestra probabilísticamente y con ella hacer la encuesta entre los consumidores dado que es menos costoso y se lleva menos tiempo realizarla que si entrevistáramos a todos los consumidores.

Paso 2.- La muestra debe de ser aleatoria, i.e., determinada matemáticamente para sustentar la confiabilidad o representatividad de sus resultados (como estimadores del universo) con base en criterios estadísticos y probabilísticos.

Paso 3.- Cuando no se tienen referencias estadísticas (medias, varianzas, etc.) se recurre a fórmulas prácticas como las siguientes (posteriormente se expondrán casos en que se dispone de antecedentes estadísticos):
n= σ al cuadrado/ α*ε al cuadrado, donde
n= tamaño de la muestra;
σ= desviación estándar al cuadrado;
α= probabilidad de que no se obtenga el error permitido o diferencia máxima entre la media muestral y µ.
ε= el error permitido o diferencia máxima entre la media muestral y µ.

Paso 4.- Al respecto, el investigador define el ε y α. Así, si con ε= 3 y si α= 5% , con σ cuadrada= 10, y si deseamos estimar la media aritmética o valor promedio del interés de los consumidores por cierto producto o servicio de la empresa que representamos, el tamaño de la muestra se obtiene aplicando la fórmula de arriba.

Observaciones: *Cuando no se conoce ningún dato* que sirva de referencia para calcular probabilísticamente la muestra se aplica la CONSIDERACION 2, descrita en el punto VI.5 DETERMINACION DEL TAMAÑO DE LA MUESTRA.
Con los datos anteriores n=10/(0.05)*9=10/0.45=22

Comentarios: En este caso se conoce la varianza poblacional; cuando no es así, se puede utilizar en su lugar la varianza de la muestra obtenida en la encuesta anterior. También, si sólo se conoce el rango de oscilación de los valores de los datos, dicho rango elevado al cuadrado se divide entre 6. Ejemplo: si los datos extremos de la serie estadística son 30 y 90, entonces la varianza es igual a (90-30) al cuadrado entre 6 es decir, 3,600/6= 600.
Como puede observarse, se intenta que el investigador disponga de datos en cualquier circunstancia para calcular matemáticamente el tamaño de la muestra, dados los beneficios estadísticos que reporta este procedimiento.
Paso 6.- Elaborar el Cuestionario, probarlo, llenarlo realizando la encuesta para obtener los resultados muestrales deseados, con base en las sugerencias que se describen en la “Red de actividades para realizar una encuesta” , mismas que se presentan en los puntos VI.6.4 y VI. 6.5.

- **MUESTREO POR ÁREAS, COMBINADO CON EL SIMPLE ALEATORIO Y EL ESTRATIFICADO.**

Por ejemplo: Considérese el siguiente diseño muestral hecho para captar las características del gasto familiar en consumo en 2002 y 2003.

Se diseñó una muestra probabilística multietápica del país que fue dividido en áreas. En un muestreo multietápico, cada persona (y familia) en el universo bajo estudio, tiene una probabilidad de ser incluida en la muestra, la cual está asociada con las probabilidades de selección de la unidad de muestreo en la cual se localiza la persona, en cada una de las etapas.

Lo primero que se hizo fue seleccionar con números aleatorios a las unidades de muestreo de la primera etapa que eran de dos tipos; áreas urbanas y áreas rurales. En la segunda etapa, con números aleatorios se seleccionaron áreas más pequeñas o manzanas dentro de las unidades de la primera etapa, seleccionadas previamente. La tercera etapa consistió en la división de las manzanas en áreas más pequeñas llamadas segmentos; con números aleatorios se seleccionaron los segmentos donde el entrevistador debía tener la información de cada una de las familias que lo integraban. Finalmente dentro de cada familia todos los adultos más uno de cada tres adolescentes seleccionados aleatoriamente, contestaron el cuestionario.

En este caso particular el modelo muestral comprendió tres etapas. La **estratificación** en el muestreo por áreas se hace generalmente en la primera etapa (es decir, las áreas se integran en estratos), ya que a partir de ella la población debe dividirse en forma tal, que se asegure la representatividad de los estratos. En el ejemplo que se ilustra, todas las unidades de muestreo de la primera etapa, áreas urbanas y rurales, fueron agrupadas en estratos de acuerdo con ciertos criterios para minimizar la variabilidad dentro de los estratos. Los criterios usados fueron flexibles ya que el propósito principal era obtener hasta donde fuera posible homogeneidad en las unidades de muestreo en la primera etapa de cada una de los estratos, así como la integración de estos últimos con un número aproximadamente igual de familias. Se seleccionaron automáticamente 14 áreas urbanas, porque contenían un número de familias mayor que el establecido por estrato.

Del resto de las áreas urbanas, se seleccionó una de cada estrato, con probabilidad proporcional a su tamaño. Similarmente en los estratos rurales, un pueblo o área fue seleccionado con probabilidad proporcional a su tamaño.

En total, se seleccionaron 103 unidades de la primera etapa, conteniendo 191 poblaciones. De las 103 unidades de la primera etapa; 49 eran urbanas y 54 rurales.

Una vez que se han diseñado las áreas y agrupado en estratos, en cada estrato se seleccionan ciertas áreas usando algún criterio, generalmente se aplica el llamado “probabilidad proporcional al tamaño", con el cual cada área tiene una probabilidad (proporcional) de ser seleccionada de acuerdo a su tamaño o significación dentro del estrato. Por ejemplo: Supongamos que deseamos seleccionar con probabilidad proporcional a su tamaño una de las siguientes cinco ciudades que integran un estrato:

		Población	Dígitos			
Ciudad	Población	Acumulada	Aleatorios	Probabilidad		
		(en miles)				
A	100,000	100	01 - 10	10	÷	35
B	40,000	140	11 - 14	4	÷	35
C	60,000	200	15 - 20	6	÷	35
D	70,000	270	21 - 27	7	÷	35
E	80,000	350	28 - 35	8	÷	35
Total estrato 350,000				35	÷	35

Un procedimiento es la selección de un número aleatorio formado por dos dígitos de cualquier tabla de números aleatorios, y luego seleccionar la ciudad cuyo rango de dígitos incluye los números aleatorio. Si el número aleatorio es mayor que 35, nuevamente se seleccionan otros números hasta obtener uno que sea igual a 35 o menos.

Por ejemplo: Si el número aleatorio es el número 22 se selecciona la ciudad D como la muestra del estrato, porque de acuerdo con la penúltima columna del cuadro anterior, el 22 es uno de los siete dígito que representan la ciudad D: Si fuera 06, la muestra contendría la ciudad A.

En esencia, se sigue el mismo procedimiento para seleccionar las manzanas de la segunda y las familias de la tercera etapa del muestreo por áreas, ya que por lo general no se requieren estratificaciones adicionales. Así, si la ciudad A es seleccionada en la muestra podría dividirse en manzanas y seleccionarse con probabilidad proporcional unas cuantas de estas con la ayuda de la tabla de los números aleatorios.
Una vez seleccionadas las manzanas, las familias se listarán en cada manzana y el número requerido de ellas se obtendría usando una vez más la tabla de números aleatorios.

Obsérvese que en poblaciones grandes y dispersas este procedimiento resulta ventajoso no sólo en la fase de la entrevista, sino también en la fase de preparación del marco muestral, ya que las definiciones y listados de las familias solo se hacen para las unidades de la primera etapa que caen en la muestra y los listados de familias se requieren solamente de aquéllas manzanas consideradas en la muestra.

MUESTREO POR RACIMOS O CONGLOMERADOS

Este método, que es en esencia una extensión del muestreo por áreas, consiste en la aplicación/uso de las *últimas unidades del muestre*o en localidades adyacentes en lugar de permitir su dispersión en todas las áreas que comprenden la muestra.

Por ejemplo: Una muestra de 300 familias podría obtenerse seleccionando 60 grupos de 5 manzanas en lugar de seleccionar individualmente a 300 familias.

Esta concentración de las unidades de muestreo reduce considerablemente el tiempo y dinero estimados para el llenado del cuestionario, por lo que se aconseja cuando el entrevistador tenga que cubrir una gran área como en el caso del muestreo en áreas rurales. Sin embargo con este se pierde cierta eficiencia/representatividad de la muestra de las características del universo.
Esta pérdida se deriva de la tendencia que tienen por vivir como vecinos las personas con iguales características, actitudes o aún hábitos de consumo. Así, una persona de altos ingresos es más probable que este al lado de otra de igual nivel; y no de una de bajos ingresos, lo que ocasiona que las unidades de muestreo en lugar de ser independientes estén correlacionadas. Mientras más alta sea la correlación positiva, menor será la eficiencia del método por racimos en la representación de las características del universo; en consecuencia, la ineficiencia resulta de la reducción en la precisión de los estimadores muestrales, dado que representarán sólo a una parte del universo.

MUESTREO REPLICADO

Hasta el momento, se han ilustrado métodos que requieren la selección de una sola muestra de la población. Un procedimiento alternativo es dividir la muestra en un número igual de sub-muestras y seleccionar cada una de las sub-muestras de la población como si cada una de ellas fuera la única muestra a seleccionar.

La muestra total, consiste en un número de sub-muestras replicadas, cada una de ellas tratando de proporcionar en su área de influencia una imagen completa del universo. Si se desean entrevistar 400 personas en un área de 10 000 personas, cada: 25, ... (10 000 entre 400) sería entrevistado comenzando con un número aleatorio entre 01 y 25.

Si se decide seleccionar 5 en lugar de una muestra cuyo tamaño total sea de 400 personas, cada una de las cinco sub-muestras deberá contener 80 unidades de muestreo. Para ello se puede dividir a la población en 125, (10 000 entre 80 = 125). Son así iguales cada una conteniendo 80 unidades de muestreo; luego se seleccionan 5 números aleatorios entre 01 y 125 que se consideran, cada uno como punto de arranque o primer unidad de muestreo que faltan en cada sub-muestra, se seleccionan progresivamente cada 125 familias. El resultado, son 5 sub-muestras replicadas o interpenetrantes con 80 unidades cada una, que agregadas suman una muestra con 400 unidades de muestreo.

DEFINICIONES BÁSICAS *(Sánchez et al, 1974)*

Error de muestreo:

Sea μ el valor de un parámetro de la población que se estudia mediante el muestreo, y $\bar{x}$ una función definida mediante la muestra, que estima el valor de μ.

Error de muestreo $= \left| « - \bar{x} \right|$ que debe ser menor o igual al máximo error de variación permitido ε| μ |; es decir ε| μ | ≥ | μ - $\bar{x}$ | .; como no conocemos μ

decimos: ε| $\bar{x}$ | ≥ | μ - $\bar{x}$ |

LÍMITES DE CONFIANZA:

Lo antes dicho se basa en el hecho de que cuando no se conocen los parámetros (μ y σ) de la población se pueden estimar recurriendo a muestras que permiten calcular intervalos dentro de los cuales puede estar contenido el valor de los parámetros. Estos intervalos se llaman intervalos de confianza y sus extremos se llaman límites de confianza.

El grado de confianza de que el parámetro está contenido en el intervalo se determina por el número de errores estándar a los cuales les corresponde un área bajo la curva que se denomina "coeficiente de confianza" (ε épsilon). Al riesgo de que el valor estimado de « no se encuentre dentro del intervalo de confianza construido alrededor de la media de la

muestra, se le llama nivel de significación (α) y es el área o probabilidad complementaria del coeficiente de confianza.
Así $\varepsilon = 1 - \alpha$ ó $\varepsilon + \alpha = 1$ = área bajo la curva.

De esta manera el intervalo de confianza se determina con:
Límites de confianza $= \bar{x} \pm Z \sim \pm_{\bar{x}}$

Donde: $\bar{x}$ = Media muestral;
Zα = Valor especifico de Z en la tabla, asociado con determinado valor de α y ε;
Error estándar para una población infinita.

$$\pm_{\bar{x}} = \frac{\pm}{\sqrt{n}}$$

n = Tamaño de la muestra;
σ = Desviación estándar de la población.

- **DISTRIBUCIÓN DE MUESTREO[(10)]**

Distribución de Muestras (de medias y de proporciones)
Por analogía, la distribución de muestreo que se deriva del universo, con determinado tamaño de muestra n y $\pm_{\bar{x}}$, tendrá

$«\bar{X} = E(\bar{X})$ y una varianza $(\bar{X}) = \frac{\pm^2}{n}$ para una población infinita y Var $(\bar{X}) = \frac{\pm^2}{n} \frac{N-n}{N-1}$ para una población finita donde

$\pm^2$ = Varianza del universo. La varianza de $\bar{X}$ se representa con $\pm^2_{\bar{X}}$, cuya raíz cuadrada $\pm_{\bar{x}}$ se denomina ERROR ESTÁNDAR para distinguirla de $\pm$ = Desviación estándar del universo o raíz cuadrada de $\pm^2$. Luego en una distribución de muestreo

$$«\bar{X} = E(\bar{X}) \text{ y } \pm\bar{x} = \frac{\pm}{\sqrt{n}}$$

Ejemplo: Supóngase la población N =3 con los términos

(xi) : 1, 2 y 3.

Su; $« = \frac{1+2+3}{3} = \frac{\sum x_i}{N} = 2$

Su; $\pm = \sqrt{\frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3}} = \sqrt{\frac{\sum (xi - «)^2}{N}} = \sqrt{\frac{2}{3}} = 0.81$

CUYOS VALORES SON FIJOS

Si tomamos muestras de tamaño 2, esto es n = 2 de N = 3 sin reemplazo, habrá

$$\begin{bmatrix} \boldsymbol{N} \\ \boldsymbol{n} \end{bmatrix} = \frac{\boldsymbol{N}!}{(\boldsymbol{N}-\boldsymbol{n})!\boldsymbol{n}!} = \frac{3*2*1}{(3-2)!2!} = \frac{3*2*1}{1!(2*1)} = \frac{6}{2} = 3$$

Interpretación: Hay tres muestras de tamaño 2, cuya composición de cada una es: 1, 2; 1, 3; y 3,2.

Estandarizando la nueva variable aleatoria $\boldsymbol{\overline{X}}$, tendremos:

Muestra	$\overline{X}_i$	$\overline{X}_{i-} - \text{«}$	$Z_i = \frac{\overline{X}_i - \text{«}}{\pm_{\overline{X}}}$	Ordenada Y_i	Área bajo la curva
1, 2	1.5	-0.5	-1.25	0.18265	0.394
1, 3	2.0	0.0	0.0	0.39894	0
2,3	2.5	0.5	+1.25	0.18265	0.394
		0	0		

$$\pm_{\overline{X}} = \frac{\pm}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{0.81}{\sqrt{2}}\sqrt{\frac{3-2}{3-1}} = \frac{0.81}{1.41}\sqrt{\frac{1}{2}}$$

$\pm_{\overline{X}}$ =0.56 (0.7) = 0.4

que nos sirve para graficar los valores estandarizados de las tres $\boldsymbol{\overline{X}}$: 1.5, 2.0 y 2.5, obteniéndose:

Obsérvese que aún cuando N = 3, es demasiado pequeña, *esta distribución tiende a la normal por el teorema del límite central*. Donde:
Xi : valores originales.
Zi : valores originales ahora expresados en unidades de desviación estándar
« : Media del universo.

$E(\overline{X})$: Esperanza matemática de las $\boldsymbol{\overline{X}}$

Luego usando la distribución de muestreo vemos que hay tres medias muéstrales (1.5, 2.0 y 2.5) llamadas "ESTADÍSTICAS", que cada una de ellas puede estimar el valor verdadero del parámetro «x que generalmente se desconoce su valor en la vida real, el cual podemos estimar que está en el rango $\left|\text{«} - \overline{X}\right|$ = error de muestreo, con cierto grado de confianza.

El ***error de muestreo*** o precisión en la estimación se mide y se calcula con las fórmulas del ***error estándar*** (en términos de probabilidad) de la media o de la proporción según sea el caso.

Así supongamos que deseamos estimar el valor de $\text{«}x$, para ello supongamos que seleccionamos aleatoriamente la muestra A, que está compuesta por las unidades de muestreo 1 y 2 y por consiguiente tiene una media aritmética ($\overline{x}$) = 1.5 y una desviación estándar de (s) = 0.5.

Muestra	Composición	Media de la Muestra $\overline{X}$ y	Desviación estándar de la muestra (s)
A	1,2	1.5	$\sqrt{\frac{(1-1.5)^2+(2-1.5)^2}{2}} = \sqrt{\frac{0.5}{2}} = 0.5$
B	1,3	2.0	$\sqrt{\frac{(1-2)^2+(3-2)^2}{2}} = \sqrt{\frac{2}{2}} = 1.0$
C	2,3	2.5	$\sqrt{\frac{(2-2.5)^2+(3-2.5)^2}{2}} = \sqrt{\frac{0.5}{2}} = 0.5$

La media ($\overline{X}$) y desviación estándar (**S**) de las muestras difieren según la muestra elegida, pero;

$$E(\overline{X}_i) = \frac{6}{3} = 2 = \text{«}x = \text{«}\overline{x} = \frac{\sum \overline{X}_i}{N}$$

Se pueden generar distintas distribuciones a partir del cálculo de la muestra con o sin reemplazo.

Cuando la selección es con reemplazo se usa la fórmula $N^n = 3^2 = 9$
Interpretación: hay 9 muestras de tamaño 2, cuya composición es:

Muestra	Composición	Media de la muestra (Xi)	P (xi)
A	1,1	1.0	1 ÷ 9
B	1,2	1.5	1

			÷ 9
C	1,3	2.0	1 ÷ 9
D	2,1	1.5	1 ÷ 9
E	2,2	2.0	1 ÷ 9
F	2,3	2.5	1 ÷ 9
G	3,1	2.0	1 ÷ 9
H	3,2	2.5	1 ÷ 9
I	3,3	3.0	1 ÷ 9
		18.0	9 ÷

			9

$$\text{«}\bar{x} = \frac{\sum \bar{x}_i}{N} = \frac{18}{9} = 2 = E(\bar{x})$$

$$\text{«}\bar{x} = E(\bar{x})P(\bar{x}) = \frac{1/9 + 1.5/9 + ... + 2.5/9 + 3/9}{9} = \frac{18}{9} = 2$$

Las distribuciones de muestras más importantes son:

a) De muestreo: Medias, que se obtienen con: Teorema de Límite Central y Ley de los Grandes Números.
b) De proporciones. que también se obtienen con el Teorema de Límite Central y la Ley de los Grandes Números.

Por otra parte cuando el tamaño de la muestra (n) y la población (N) son grandes, el número de muestras es muy grande; por lo que se recomienda simplificar el procedimiento de obtención de muestras probabilísticas.

- **Teorema de límite central**

Con este objeto, se usa el **teorema del Límite Central** para demostrar que se puede utilizar la media de la muestra para representar la media de la población.

Su aplicación da sustento a la teoría de la estimación, es decir, a la inferencia estadística porque con él se puede inferir, a partir de la media muestral el valor y comportamiento de la media poblacional. En general, a partir de los resultados de la muestra, las características de los datos del universo estadístico. Ello es así porque es común usar la distribución de probabilidad normal como una aproximación a la distribución de muestras (sean sus “estadísticas” medias o proporciones) cuando el tamaño de la muestra es mayor de 30 datos, en virtud de que la distribución de probabilidad normal hace factible estimar valores poblacionales a partir de valores muestrales.

El teorema del Límite Central, establece que si una población es normal, con media y desviación estándar, µx y σx, entonces si tomamos muestras de tamaño n y a estas les calculamos sus medias aritméticas, la nueva distribución constituida por las medias de las muestras, es una distribución muestral, normal con:

$\text{«}\bar{x} = E(\bar{x})$ y $\pm_{\bar{x}} = \frac{\pm x}{\sqrt{n}}$ para una población infinita

- **Ley de los grandes números**

La ley de los Grandes Números establece que si una población tiene µx y σx independientemente de que sea o no normal; si el tamaño de la muestra n crece, entonces la distribución que resulta de las medias muestrales se aproximan a la normal con $\text{«}\bar{x}$ y $\pm_{\bar{x}}$.

Para demostrar lo anterior y trabajando con los datos conocidos:

Medias de las muestras ($\bar{X}$)	**P($\bar{X}$)**
1.5	1 ÷ 3
2.0	1 ÷ 3
2.5	1 ÷ 3

$$E(\bar{x}) = \frac{1.5}{3} + \frac{2.0}{3} + \frac{2.5}{3} = \frac{6}{3} = 2 = \text{«}\bar{x} = \text{«}$$

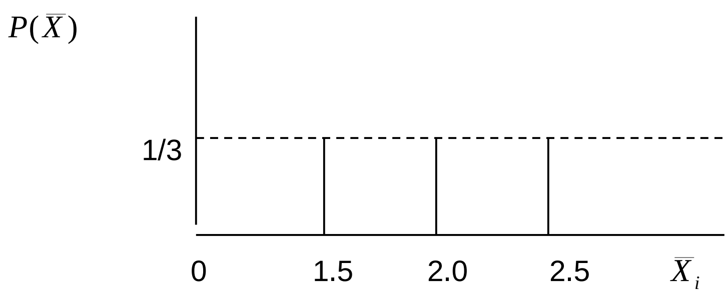
$$\pm\bar{x} = \sqrt{\frac{(1.5-2)^2 + (2-2)^2 + (2.5-2)^2}{3}} = \sqrt{\frac{0.50}{3}} = \sqrt{0.17} = 0.4$$

También $\pm\bar{x}$ se obtiene con $\pm\bar{x} = \frac{\pm x}{\sqrt{n}}$ cuando n es muy grande

$$\pm\bar{x} = \frac{\pm x}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{0.81}{\sqrt{2}}\sqrt{\frac{3-2}{3-1}} = \frac{0.81}{1.41}\sqrt{\frac{1}{2}} = \frac{0.81}{1.41}(0.7) = (0.57)(0.7) = 0.4$$

$$\pm\bar{x} = 0.4$$

Gráfica de la nueva distribución de muestreo con «$\bar{x}$ = 2 y $\pm\bar{x}$ =0.4



valores originales, sin estandarización

Si se desea calcular el intervalo de confianza dentro del cual se halle contenido el valor de μ, para calcular el intervalo de confianza el investigador determina el grado de confianza, (γ) deseado en la estimación. El grado de confianza lo determina el número de errores estándar deseado, que en términos de probabilidad, a su vez determina el error de muestreo. Así, para la primera muestra sabemos que:

n = 2

$\bar{x}$ = 1.5

s = 0.5

con γ= 95% probabilidad (área bajo la curva) de que μx se halle en el intervalo $\overline{X} \pm Z{\sim}\pm\overline{x}$; donde α= 5% = probabilidad de que no sea así, se denomina nivel de significación.

Derivado de lo anterior diremos que a un γ = 95% le corresponden 1.96 errores estándar = 1.96 $\pm\overline{x} = Z{\sim}\pm\overline{x}$.

Así $\overline{X} \pm Z{\sim}\pm\overline{x}$ y como; $\pm\overline{x} = \dfrac{s}{\sqrt{n}} = \dfrac{0.5}{\sqrt{2}} = \dfrac{0.5}{1.41} = 0.35$

por lo tanto 1.5 ± 1.96 (0.35)

1.5 ± 0.70

luego el límite inferior del intervalo es 0.80 = 1.50 - 70 y el límite superior del intervalo es 2.20 = 1.50 + 0.70.

Interpretación: Hay una probabilidad del 95% que el valor μx se halle en el intervalo de 0.80 a 2.20.

Generalizando, si la muestra seleccionada hubiera sido la B o la C, tendríamos:

B	**C**
$\overline{X}$ = 2	$\overline{X}$ = 2.5
s = 1.0	s = 0.5
n = 2	n = 2

Para B $\pm_{\overline{x}} = \dfrac{s}{\sqrt{n}} = \dfrac{1}{\sqrt{2}} = \dfrac{1}{1.41} 0.70;$

$\overline{X} \pm Z{\sim}\pm\overline{X}$

$2 \pm 1.96(0.70)$

2 ± 1.37

Intervalo: De 0.63 a 3.37

para C, $\pm_{\overline{x}} = \dfrac{s}{\sqrt{n}} = \dfrac{0.5}{\sqrt{2}} = \dfrac{0.5}{1.41} 0.35$

$\overline{X} \pm Z{\sim}\pm\overline{X}$

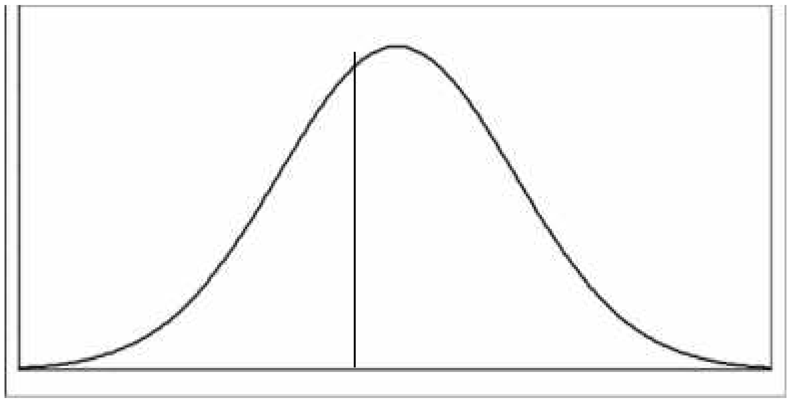
$2.5 \pm 1.96(0.35)$

2.5 ± 0.70

Intervalo: De 1.80 a 3.20

Conclusión: En los tres casos el valor de x = 2 se halla contenido con una seguridad, probabilidad o confianza del 95% y con un riesgo de = 5% de que no sea así, en los intervalos antes calculados.

Gráficamente.:



$\bar{x} - Z_{\pm_x}$ $«_x$ $\bar{x} + Z_{\pm_x}$

A: 0.80=1.5-0.70 1.5 1.5+0.70=2.20
B: 0.63=2.0-1.37 2.0 2.0+1.37=3.35
C: 1.80=2.5-0.70 2.5 2.5+0.70=3.25

Si conectamos estos resultados con la definición básica de que el error de muestreo $(\bar{x} - «)$ se determina con el error estándar de la media , en términos de probabilidad, $\pm\bar{X}$, y con la situación ideal de que siempre esperamos que el error de muestreo sea igual o menor al error permitido (e), observamos que:

1.- con la muestra 1: e $|«\, X| \geq |\bar{x} - «|$ ya que e=0.70 $\geq |1.5 - 2.0|$

2.- con la muestra 2: tenemos e= 1.37 $\geq |2 - 2|$ y

3.- con la muestra 3: vemos que e= 0.70 $\geq |2.5 - 2.0|$,

en los tres casos es satisfactorio ver que el error de muestreo es inferior al error permitido.

Nuevo ejemplo; ahora supongamos que £ = 50% y Zα = 0.68.
luego α = 50%

						Limites		Contiene a
Muestra	$\bar{X}$	s	$\sigma\bar{X}$	Zα	Zασ	Inferior	Superior	μx

					$\overline{X}$			
A	1.5	0.5	0.35	0.68	0.238	1.262	1.738	No
B	2.0	1.0	0.70	0.68	0.476	1.524	2.476	Si
C	2.5	0.5	0.35	0.68	0.238	2.262	2.738	No

La muestra A y C no contienen a μx porque el grado de confianza ξ es muy bajo; es decir hay menos área sobre la curva que ocasiona una Zα muy baja que al ser combinada en Zασ$\overline{X}$ originan un intervalo más pequeño en torno a $\overline{X}$, en la fórmula $\overline{X}$ ± Zασ$\overline{X}$, con lo que aumentan la probabilidad α, de que $\overline{X}$ no represente a μx. Estos resultados se corroboran con el siguiente análisis:

Con la muestra 1: e=0.238$\leq |1.5-2.0|$, por eso el intervalo de confianza no contiene a la media poblacional;

Con la muestra 2: e= 0.476$\geq |2.0-2.0|$, por eso contiene a la media poblacional y con la muestra 3: e= 0.238$\leq |2.5-2.0|$, por eso no contiene a la media poblacional.

- **ERROR PERMITIDO Y ERROR DE MUESTREO.**

De lo anterior podemos decir que e = error permitido = Zασ$\overline{X}$.

Se dice que es el error permitido;   y n condicionan los valores de Zα y de σ$\overline{X}$.

Así, como: e = Zα $\pm\overline{x} = \dfrac{\overline{x} - «\overline{x}}{\pm\overline{x}} * \dfrac{\pm x}{\sqrt{n}} = \dfrac{\overline{x} - «\overline{x}}{\pm\overline{x}} \pm\overline{x} = \overline{x} - «x$

$e = |\overline{x} - «x|$ = Error de muestreo; también: $e|«x| =$ error permitido.

Idealmente siempre queremos que $e|«x| \geq |\overline{x} - «x|$. Observe que ambos requieren del error estándar $(\pm\overline{x})$ para su cálculo.

Por otra parte mostrando los valores de mayor uso de Zα, ξ y α, de la ecuación (1) tenemos: $(\pm\overline{x})$ para su cálculo.

$\overline{X}$ - Zα $\pm\overline{x}$; límite inferior del intervalo

$\overline{X}$ + Zα $\pm\overline{x}$; límite superior del intervalo

Zα	1.00	1.96	2.00	3.00
ξ	0.68	0.95	0.955	0.997
α	0.32	0.05	0.045	0.003

Ejemplo: Se desea conocer el ingreso medio de los trabajadores de la Cía. PEPSI COLA con el fin de estudiar las condiciones de trabajo y en su caso pedir mejorías en la revisión del

Contrato Colectivo de Trabajo. Para ello seleccionamos una muestra aleatoria de 49 trabajadores cuyo ingreso medio es de $ 5,500 /mes.

Estudios previos realizados por la Facultad de Contaduría y Administración revelan que la ± del universo es de $ 700/ mes. Con α = 5%, determinar el intervalo de confianza dentro del cual se halla ingreso medio de los trabajadores.

n = 49 $\quad\quad \bar{X} \pm Z\alpha\sigma\bar{X}$

sustituyendo

± = 700/mes $\quad\quad 5500 \pm 1.96(100)$

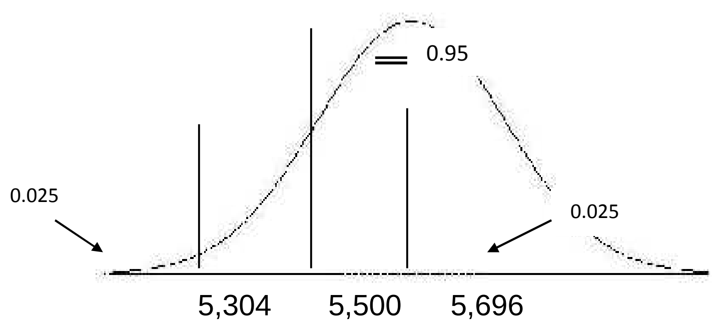
$\bar{X}$ = 5500/mes $\quad\quad 5500 \pm 196$

α = 5 %

Zα = ± 1.96

$\pm\bar{x} = \frac{\pm x}{\sqrt{n}} = \frac{700}{\sqrt{49}}$; por lo tanto $\pm\bar{x}$ = 100

Limites de confianza = 5,500 ± 196
Intervalo de confianza: 5,304 a 5,696
Donde el límite inferior = 5,304
El límite superior = 5,696
INTERPRETACIÓN: El ingreso medio x de los trabajadores de la PEPSI se halla entre los $5,304 y $ 5,696 con una probabilidad o seguridad del 95%.



$\bar{x} - Z_{\sim}\pm_{\bar{x}} \; x \; \bar{x} + Z_{\sim}\pm_{\bar{x}}$

En este caso se estima µx con la variable aleatoria asociada x mediante $\bar{X}$ proveniente de n = 49 con α = 5% y un 1- α = 95% que les corresponde una Zα = 1.96 = Número de desviaciones estándar y $\pm\bar{x}$ = 100,
tal que:

$P(\bar{x} - Z{\sim}\pm\bar{x} \leq «x \leq \bar{x} + Z{\sim}\pm\bar{x}) = 1 - {\sim} = 95\%$; en otras palabras digamos que $e|«x| = 6\%$ con $|«x|$ tenemos:

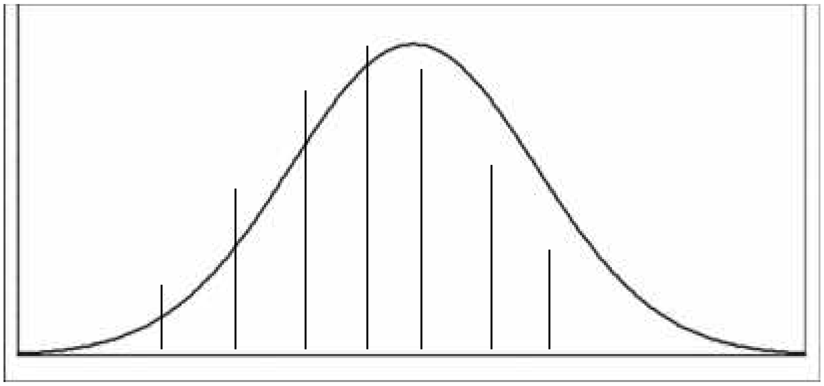
$$P|«x - \bar{x}| \geq e|«x| = 0.06|« \, \chi \,| = 1 - ¶ = 1 - 0.95 = 5\% = {\sim}$$

Ello significa que el error en la estimación del valor de µx en valores absolutos es:
| error en la estimación de µx | = Zx$±\bar{x}$, por lo que

error máximo permitido = error en la estimación de $e|«\,x|$

Derivado de lo anterior se puede escribir e = Zα $±\bar{x}$

Gráficamente dichas relaciones se ven así:



$«x = E(\bar{x})$

$\bar{x} - 1±_{\bar{x}}\ \ \bar{x} + 1±_{\bar{x}}$

$\bar{x} - 2±_{\bar{x}}\ \ \bar{x} + 2±_{\bar{x}}$

$\bar{x} - 3±_{\bar{x}}\ \ \bar{x} + 3±_{\bar{x}}$

donde $±\bar{x} = \frac{±}{\sqrt{n}}$ para una población infinita

$±\bar{x} = \frac{±}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ para una población finita

- **DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA (n)**

Determinación probabilística del tamaño de la muestra

Para que sus resultados sean representativos estadísticamente y conozcamos a priori el “ error de muestreo” *es necesario determinar* arbitrariamente o con conocimiento de causa: 1.- el error permitido (ε) cuyo valor es la diferencia máxima que el investigador acepta que exista entre el estimador muestral (la media aritmética de la muestra y el parámetro poblacional correspondiente, la media de la población: µ y 2.- el nivel de confianza (ξ) o probabilidad con que se asegura lo anterior, tal que ξ= 1-α, donde α es la probabilidad de que no se cumpla lo esperado (diferencia máxima entre el valor muestral y el poblacional) ; así si por ejemplo si conocemos la desviación estándar de la población (σ) podemos determinar el tamaño de la muestra (n) así:

n= σ al cuadrado/α*ε al cuadrado

Ejemplo. Si conocemos σ cuadrada = 12 y deseamos que ε = 3, es decir que la media aritmética de la muestra no se aleje en más de 3 puntos como máximo de μ, con una ξ del 95%, entonces α= 5% tendremos que
n= 12/ (0.05)*(9)= 27

En general, se puede o*btener el tamaño de la muestra con literales previamente definidas de la siguiente manera:*

Sabemos que:

$e = Z\sim \pm\bar{x}$ como $\pm\bar{x} = \frac{\pm x}{\sqrt{n}}$ para una población infinita

$$e = Z\sim \frac{\pm x}{\sqrt{n}}$$

$$\sqrt{n} = \frac{\pm x}{e} Z\sim$$

$n = \left[\frac{\pm x}{e} Z\sim\right]^2 = \frac{Z\sim^2 \pm^2 x}{e^2}$ para una población infinita.

$n = \frac{Z^2 \sim \pm_x^2 N}{e^2 N - e^2 + Z\sim^2 \pm_x^2}$ para una población finita.

Ejemplo: En una población infinita ¿qué, tamaño de muestra será necesario para producir un intervalo de confianza del 90% en que está la media de la población verdadera, con un error permitido máximo de diferencia entre la media muestral y la poblacional de 1.0 en cualquier sentido si la desviación estándar de la población es 10.00?

Solución: Sabemos que $\pm_x = 10.0$

e = 1.0

Z α = ± 1.65

Para α = 10 %

$$n = \left[\frac{\pm x}{e} Z\sim\right]^2 = \left[1.65 \frac{10.0}{1.0}\right]^2 = (1.65)^2 = 272.2$$

Con este tamaño de muestra de 272.2 aseguramos que el error de muestreo = $\left|\bar{x} - «_x\right| \leq error\ permitido = Z_\sim \pm_{\bar{x}}$ donde $\pm_{\bar{x}} = \frac{\pm_x}{\sqrt{n}}$

Así, con $Z_{\sim} = 1.65$ y $\pm_{\bar{x}} = \frac{10}{\sqrt{272.25}}$

$e = 1.65(10/\sqrt{272.25}) = \frac{16.5}{16.5} = 1 = error \quad permitido = errordemuestreo$, lo cual es muy aceptable.

También, si sabemos que el error estándar = 10/16.5= 0.606

Luego aplicándolo al error permitido (e) en términos probabilísticos tendremos que e = 1.65 (0.606)= 1, se comprueba que el error de muestreo se mide con el error estándar en términos probabilísticos.

CONSIDERACIONES:

1. Hay ocasiones en que conocemos N, en ese caso $n = \frac{N}{Ne^2 + 1}$

Ejemplo: Con N= 603 y e= 5%

Tenemos $n = \frac{603}{1 + 603(0.05)^2} = \frac{603}{2.5075} = 240.47$

2. Cuando no conocemos nada $n = \frac{1}{e^2}$ digamos si e = 5 %

$$n = \frac{1}{(0.05)^2} = \frac{1}{0.0025} = 400$$

3. Trabajando con proporciones o atributos diremos que en el muestreo simple aleatorio: cada elemento tiene la misma probabilidad de ser seleccionado y, por ejemplo con n = 300 y α = 5%, ξ = 95% $Z\alpha$ = 1.96, el error permitido (e) o margen de error permitido para p = 0.5 = q será igual a:

$$e = \sqrt{\frac{pq}{n}} * Z\sim \; = \sqrt{\frac{(0.5)(0.5)}{300}} * 1.96$$

$$e = \pm p * Z\sim \; = 5\%$$

- **EVALUACIÓN DEL TAMAÑO DE LA MUESTRA [23]**

Partiendo de $n = \frac{Z^2 \pm^2}{e^2}$ donde e: es el error mínimo permitido, que lo determina el investigador, ya que sólo él está en condiciones de fijarlo para aceptar su resultado muestral. Por ejemplo puede especificar que si la media obtenida de la muestra es $ 6 mayor o menor que la media verdadera (poblacional), considerará que el estimador $\bar{x}$ obtenido mediante la muestra es satisfactorio. Por lo tanto e = $6, y el intervalo de confianza es $\bar{x} \pm$ $ 6.

$Z\alpha$ se establece mediante el nivel de confianza del intervalo; por ejemplo si el investigador desea que el resultado de la estimación sea $\xi = 99.73\%$ prácticamente seguro, ξ= 99.73%, de

[23] Tomado del Profesor Stephen P. Shao.

que la media estimada de la población con base en la muestra esté dentro del recorrido de la verdadera media de la población ± \$ 6 ó µx ± \$6, el valor de Zα es 3.
Así, una vez que tenemos el tamaño de la muestra, el resultado de la muestra debe ser evaluado. Esto puede ser hecho encontrando el ERROR ESTÁNDAR DE LA MEDIA $S\bar{x}$, de acuerdo con la desviación estándar de la muestra $\hat{s}$.
45

Si el producto de Zα $S\bar{x}$ es menor que el error de muestreo especificado (permitido), la estimación de la muestra es considerada satisfactoria. Si el producto es mayor, el tamaño de la muestra deberá ser revisado e incrementado.
Ejemplo: El Gerente de una estación de servicio desea muestrear las notas de venta a fin de encontrar la cantidad (media) promedia por venta durante un período dado.
Para ello indica que 1) el máximo error muestral (permitido) no deberá ser mayor que 20 ¢ por arriba o por abajo de la verdadera media; 2) el nivel de confianza deberá ser ξ 99.73%; y 3) la desviación estándar de la población basada en su experiencia, es estimado en 80 %. Encontrar el tamaño de la muestra adecuada con estas especificaciones.
SOLUCION:
1.-El intervalo de confianza es µx ± \$ 0.20 luego e = \$ 0.20
2.-Para ξ= 99.73 tenemos Zα = 3
3.- $n = \left[\frac{Z\sim \pm x}{e}\right]^2 = \left[\frac{3(0.80)}{0.20}\right]^2 = 12^2 = 144$ tamaño de la muestra
Ahora suponga que trabajando con esa muestra seleccionada aleatoriamente se aplica y encontramos que:
$\bar{x}$ = \$ 2.70
$\hat{s}$ = \$ 0.72
luego $S\bar{x} = \frac{\hat{S}}{\sqrt{n}} = \frac{0.72}{\sqrt{144}} = \$0.06$

Construimos el intervalo de confianza :
$\bar{x} \pm Z\sim S\bar{x} = 2.70 \pm 3(0.06) = 2.70 \pm 0.18 =$ 2.52 a 2.88

Puesto que Z $S\bar{x}$ = 0.18 = error de muestreo es menor que el error permitido e = 0.20, se acepta el tamaño de la muestra.
Sin embargo ahora supóngase que con; n = 144 $\hat{s}$ = \$ 0.84, entonces
$S\bar{x} = \frac{\hat{S}}{\sqrt{n}} = \frac{0.84}{\sqrt{144}} = 0.07$ luego:
$\bar{x} \pm Z\sim S\bar{x} = 2.70 \pm 3(0.07) = 2.70 \pm 0.21$

Como el error de muestreo calculado 0.21 es mayor que el error permitido (e = 0.20), el tamaño de la muestra se revisa como sigue, partiendo de una población infinita:

$$n = \left[\frac{Z{\sim}Sx}{e}\right]^2 = \left[\frac{3(0.84)}{0.20}\right]^2 = 158.76 = 159$$

Por lo tanto el tamaño de la muestra aumenta a 159.
Ahora bien; con Sx = 0.80
¿Cuál es el tamaño de la muestra si α = 95.45 % y Zα= 2?

$$n = \left[\frac{Z{\sim}Sx}{e}\right]^2 = \left[\frac{2(0.80)}{0.20}\right]^2 = 8^2 = 64$$

De este ejemplo numérico se deduce que el tamaño de la muestra depende significativamente de los valores que tomen e, Zα, $\pm_{x}$. En poblaciones finitas, N, es determinante.

Una vez establecida las "definiciones básicas" a continuación empezamos a aplicarlas en temas fundamentales que constituye la ESTADISTICA INDUCTIVA moderna.

Aún cuando la exposición y composición de estos temas no es fácil, yo espero que el esfuerzo didáctico que adopte le permita al lector su fácil entendimiento y manejo continuo en la solución de problemas de su empresa, principalmente, en las áreas de ventas, compras, producción, organización y finanzas.

- **PRECISIÓN - ERRORES DE MUESTREO**

Como se indicó, la confiabilidad en las estimaciones se mide por medio de los errores de muestreo, que a su vez, se determinan con las fórmulas de los errores estándar, en términos de probabilidad, es decir: $Z_{\sim} \pm_{\bar{x}}$. Con ese propósito a continuación ilustramos las fórmulas de los ERRORES ESTÁNDARES de los principales diseños muestrales, las cuales son muy importantes ya que a partir de ellas se calculan:

a) Tamaño de la muestra,
b) Límites de confianza,
c) Errores de muestreo y
d) Se prueban hipótesis.

Muestreo simple aleatorio

$\pm_{\bar{x}} = \frac{\pm}{\sqrt{n}}\sqrt{\frac{N-n}{N*n}}$; con proporciones: $\pm_{p} = \sqrt{p*q\frac{N-n}{N*n}}$

Muestreo estratificado.

$\pm_{\bar{x}} = \sqrt{\sum_{i=1}^{k} w_i^2 s_i^2 \frac{N_i - n_i}{N_i * n_i}}$; con proporciones: $\pm p = \sqrt{\sum_{i=1}^{k} w_i^2\, pq \frac{N_i - n}{N_i * n}}$

$Si^2 = pq$

donde:

i: estratos: 1,2,3,4,5,.....,K.

W_i : Proporción del estrato en la población = $\frac{N_i}{\sum N_i}$

$P_i = \frac{n_i}{n}$; n= tamaño de la muestra; n_i= muestra en el estrato i-ésimo, N_i = estrato i-ésimo.

Muestreo replicado

$\pm_{\bar{x}} = \left|\frac{\bar{X}max - \bar{X}min}{K}\right|\sqrt{\frac{K(Z-K)}{Z(K-1)}}$

donde:

$\bar{X}$ max: La media mayor en la muestra replicada.

$\bar{X}$ min: La media menor en la muestra replicada.

Z : Tamaño de cada zona

K : Número de replicaciones.

Ejemplos aplicando los errores estándar en la determinación de la precisión.

Se desea estimar con un 95 % de confianza, la proporción verdadera de familias que tienen encendida su T.V. entre las 7 y 10 de la noche. En otras palabras, se busca el intervalo alrededor de la proporción muestral.

Con N = 10,000 familias

Con n = 400 familias con televisión.

6.3.- MUESTREO SIMPLE ALEATORIO[11]

Se selecciona una muestra aleatoria y se encuentra que 280 de las 400 televisores están encendidos una o más veces en el tiempo señalado, luego el porcentaje muestral es igual a:

n_i/ n = 70% = 280/400

$$\pm_{p} = \sqrt{p * q \frac{N - n}{N * n}} =$$

$$\pm_{p} = \sqrt{0.70(0.30)\frac{10{,}000 - 400}{10{,}000(400)}} = 2.3\%$$

Por motivos prácticos decimos que en una muestra grande, dos errores estándar proporcionan el intervalo de confianza del 95.45 %, para la proporción verdadera de TV encendidas entre los 7 y 10 de la noche; la estimación del intervalo será:

70 % 2 (2.3) ó entre 65.4 % y 74.6 %.

6.3.- ESTRATIFICADO:

Estrato	N_i	Nº de entrevistas n	Nº de T.V. encendidas entre 7 y 10 hrs. n_i	$P_i = n_i / n$
1	7,000	200	160	160 ÷ 200 = 80 %
2	1,000	100	40	40 ÷ 100 = 40 %
3	2,000	100	60	60 ÷ 100 = 60 %

10,000 400 260

$$\pm_p = \sqrt{(0.70)^2(0.8)(0.2)\frac{7,000-200}{7,000*200} + (0.10)^2(0.4)(0.6)\frac{1,000-100}{1,000*100} + (0.20)^2(0.6)(0.4)\frac{2,000-100}{2,000*100}} =$$

$$\sqrt{0.0003807 + 0.0000216 + 0.0000912} = \sqrt{0.0004935}\,; \quad \pm_p = 0.022 \text{ ó } 2.2\%$$

En este caso, el intervalo es 65 % ± 2(2.2 %) ó entre 60.6% y 69.4%.

Muestreo replicado:

Aquí suponga que se usaron los 5 diseños replicado: 5 muestras de 80 personas fueron seleccionadas de la población; de cada una de las 125 zonas registradas

Replica	Nº entrevistas	T.V. encendidas	P	
1	80	59	74	%
2	80	57	71	%
3	80	61	76	%

4	80	53	66	%
5	80	62	78	%
Total	**400**	**292**	**73**	**%**

$\pm_{\bar{x}} = \left|\frac{0.78 - 0.66}{5}\right| \sqrt{\frac{5(125-5)}{125 * 4}} = 0.026$ ó 2.6%

El intervalo es 73 % ± 2 (2.6 %) ó entre 67.8 % y 78.2 %.

Vemos que el menor error estándar se obtiene en el muestreo estratificado, razón por la que siempre se recomienda usarlo.

- **DETERMINACION DEL TAMAÑO DE LA MUESTRA**

Por su importancia derivada de los ejemplos anteriores, veamos de nuevo como se obtiene el tamaño de la muestra (n) a partir de las fórmulas del error estándar, en este caso de una proporción.

Se toma una muestra para estimar entre otras cosas, la proporción de familias viendo T.V. en la tarde entre semana.

Se desea que ese estimador esté entre el 5 % del porcentaje actual con 95% de seguridad.

N = 10,000

S^2= para un porcentaje = p*q

p = .5 por seguridad, es decir, trabajando con variancia máxima.

$\pm_p$ debe ser tal que $2\pm_p$ incluyan el 95% de los estimadores de p, luego $2\pm_p = 0.05$ de aquí que $\pm_p$=0.025

De $\pm_p = \sqrt{p * q \dfrac{N-n}{N * n}}$; $\pm_p^2 = \dfrac{pqN - npq}{Nn}$ tenemos $n(N\pm_p^2 + pq) = Npq$

$n = \dfrac{Npq}{pq + N\pm_p^2}$ entonces

$n = \dfrac{(0.25)(10{,}000)}{(0.25) + 10{,}000(0.025)^2} =$ 385 familias

Vemos que el tamaño apropiado seria de 385 familias y no 400 para hacer la investigación.

De manera similar, se puede obtener los tamaños de muestra para cada uno de los modelos muestrales bajo estudio.

Resumiendo: en forma didáctica y sencilla hemos expuesto las características de los principales diseños de muestreo de mayor uso en economía. En este sentido, como una de sus aplicaciones es en la elaboración de ENCUESTAS, a continuación presentamos la relación de actividades que deben de efectuarse para hacer una encuesta.

Para afianzar el conocimiento decidí explicar algunas de las actividades, como son las siguientes:

- **DISEÑO DEL CUESTIONARIO**

El diseño de un cuestionario, involucra la consideración de un sin número de aspectos diferentes, de los cuales quizá los más importantes son:

- Los objetivos del estudio;

- La forma que debe tener;

- Si contendrá preguntas abiertas - codificación previa o posterior de las preguntas;

- La forma como se harán las preguntas;

- La organización e instrucción del cuestionario; etc.

Lo que también indudablemente determina su diseño es el tipo de datos que se desean obtener; el método usado para obtenerlos y en última instancia el uso de los resultados. Adicionalmente, podría señalarse que el diseño depende fuertemente de los antecedentes y experiencias del investigador, el tipo de entrevistadores disponibles, costo y tiempo.

Así, basándose en los formatos de la tabulación del guión de información, los rangos probables de variación tomados de las experiencias anteriores - si las hay - y las posibles respuestas, el cuestionario debe diseñarse en forma simple, fácil de seguir y si es posible atractiva.

Lo último es particularmente importante en el caso de los cuestionarios que se envían por correo, donde la decisión de los miembros de la muestra, sobre llenarlo o no, depende de la impresión que tengan sobre la apariencia del cuestionario. Al respecto, se aconseja recabar la información a través de entrevistas directas, ya que el enumerador puede inmediatamente captar los datos en forma precisa o corregirlos cuando el caso lo amerite.

El formato del cuestionario puede tener entradas múltiples o una sola; puede ser de preguntas cerradas o abiertas; las respuestas pueden estar precodificadas o no;

cuando las preguntas son abiertas, las respuestas se codificarán con base en un INSTRUCTIVO DE CODIFICACIÓN.

- **TRABAJO DE CAMPO**

Es conveniente mencionar que existen diversos métodos para la recolección de datos, de los cuales los principales son:

a) La selección de la muestra a partir de la información de los archivos de la empresa. Así, una muestra puede ser escogida sin mayor problema y al mismo tiempo los datos pueden ser obtenidos con alto grado de confianza a un costo relativamente bajo. Además de que la muestra puede mantenerse continuamente sin representar mayores cargos o esfuerzos extraordinarios;
b) Métodos de observación: La recolección de los datos por observación, es otro instrumento que indirectamente capta la información. Como la información interna, este método no requiere contacto directo con los elementos de la muestra. Estos métodos se utilizan observadores humanos y/o mecánicos, prefiriendo los primeros en casos donde haya que distinguir; por ejemplo: los adultos de los niños, o las personas por sexo.

c) Entrevistas telefónicas: cuando se puede aplicar este método resulta altamente eficiente en la recolección directa de la información. Lo anterior, se debe a que la población virtualmente esta contenida en un directorio y la selección de la muestra, se convierte en una actividad de rutina. Las entrevistas son de lo más económico -excepto cuando hayan que hacerse bastante llamadas de larga distancia- y los datos se obtienes rápidamente. Sin embargo, como los demás métodos, también tiene sus limitaciones. Obviamente no es aplicable si las entrevistas comprenden cuestiones visuales - publicidad, pruebas de interpretación, etc. - . A la vez, información altamente personal se obtiene con menos éxito por teléfono que -por ejemplo.- a través de una entrevista personal.

d) Entrevistas personales: dentro de las formas directa de obtener los datos, este método es sin lugar a dudas el más popular, por referirse a una conversación directa " frente a frente " entre un miembro de la muestra y el entrevistador. Como resultado, se puede obtener una gran variedad de información con este método, el cual es flexible en varios sentidos. Por ejemplo, los datos pueden ser registrados en grabadora o en cuestionarios.

La construcción de los cuestionarios es un arte en si; requiere numerosas precauciones para evitar respuestas sesgadas.

Desde el punto de vista de la obtención de los datos, puede decirse que existen dos formas de entrevistar: En un extremo se haya la entrevista altamente estructurada, en la cual se prepara un cuestionario formal y las preguntas se hacen bajo instrucciones precisas y el entrevistador mantiene un orden estricto para su contestación.

Esta forma, se usa generalmente para obtener una variedad de información diferente acerca de una materia, siguiendo algún orden particular. Esta forma en cierto modo, evita que la información recabada refleje sesgos debidos a juicios personales de los enumeradores.

En el otro extremo esta la entrevista carente de formalidad para la cual no se requiere un cuestionario, basta una lista de preguntas generales o temas relacionados con la información que se busca.

Dentro de estos extremos, existen varias combinaciones. El enumerador puede usar un cuestionario estructurado, pero se le permite hacer las preguntas como el quiera.

Como podrá intuirse, el enumerador es la piedra angular de una entrevista, indistintamente de la forma que adopten para entrevistar o cual sea la unidad de muestreo. Si está debidamente entrenado (a), no solamente entrevistará a un mayor número de personas, sino que los datos serán más confiables.

Parece que los mejores enumeradores son personas entre los 25 y 50 años, que tienen una evidente disposición, son inteligentes, poseen cierta cultura, son flexibles y precisos en sus hábitos de trabajo.

Indudablemente que la experiencia es útil, pero si se proporciona un buen entrenamiento puede no ser necesaria. En ciertos tipos de nuevas encuestas, la experiencia puede ser una limitante, ya que se requiere que el enumerador siga procedimientos contrarios a los acostumbrados en el pasado.

Por lo que se refiere a la organización y control del trabajo de campo, como las demás etapas requiere una programación de tiempos y actividades para asignar al personal correspondiente. Dentro de los aspectos básicos esta la fijación de las rutas de trabajo, el plan de trabajo o forma de entrevistar y la supervisión -sobre todo- cuando el grupo de trabajo es numeroso o la captación de los datos presentan dificultades.

- **CRITICA DE CUESTIONARIOS**

Los cuestionarios, codificados o no previamente, llegan a la oficina con el orden y presentación de las respuestas dadas por los enumeradores. En algunas ocasiones el trabajo se realiza de acuerdo a las instrucciones establecidas y enseguida pasa al departamento de captura, para ser procesado inmediatamente. Sin embargo, en la mayoría de los casos se requiere una crítica o revisión cuidadosa ya que:

a) Pueden traer las respuestas ilegibles;

b) El orden en que aparecen las respuestas no es el indicado;

c) Se contradicen unas respuestas con otras al compararse entre si;

d) Existen preguntas que vienen en blanco y debían haberse contestado en alguna u otra forma etc.

e) Se requiere preparar los cuestionarios para la codificación de las respuestas; y

f) Se desea verificar la autenticidad de los datos y preliminarmente comprobar ciertas hipótesis establecidas en la programación inicial de actividades, etc.

Tal que en esta etapa la información debe quedar depurada y ordenada hasta donde sea posible para su posterior transformación y vaciada en formatos previamente diseñados. En algunos casos se acostumbra usar la computadora -filtrado electrónico- para realizar esta etapa.

- **CODIFICACIÓN Y PROCESAMIENTO DE DATOS**

Una vez que los datos han sido obtenidos y revisados, deben ser procesados para hacer posible un análisis del fenómeno estudiado. Es generalmente aceptado que esta actividad es un tanto tediosa, pero también que es crítica para asegurar exactitud en los resultados.

Una tabulación hecha sin cuidado puede viciar una buena planeación y el método de obtención de los datos. Así mismo, los peligros de los sesgos a un se presentan en los procesos de preparación, clasificación y tabulación.

Esta etapa esta fuertemente ligada a la anterior, ya que, por ejemplo, la preparación consiste en la inspección de cuestionarios o cualquier otra forma usada para captar los datos, su exactitud, si están completos o no, la inspección de trabajo de campo, arreglos o eliminación de respuestas por su inconsistencia o desconfianza la clasificación o estandarización de los datos en base comunes y sobre todo su preparación para ser tabulados.

- **CLASIFICACIÓN.**

Es el arreglo de los datos en clases o categorías para ser manipulados de acuerdo con la verificación de la hipótesis de trabajo.

- **TABULACIÓN.**

La tabulación es la etapa que sucede inmediatamente después a la crítica de cuestionarios y es un conjunto de procedimientos que se adoptan para la recopilación o vaciado de los datos en cuadros. Estos últimos comprenden las diferentes relaciones que se establecen entre las variables comprendidas en el estudio, así, habrá cuadros de una sola entrada, doble entrada, etc.

Los datos pueden ser tabulados manualmente o mecánicamente. La tabulación manual se aconseja cuando las encuestas son pequeñas, existen problemas de presupuesto o no hay ninguna posibilidad de procesar los datos electrónicamente. Por el contrario, cuando la encuesta es grande, la tabulación manual, además de tardada acarrea el riesgo de cálculos erróneos por lo voluminoso de la información, aconsejándose el uso de las computadoras. Por ello, será necesario que la información sea capturada y se diseñan los programas que calcularán los datos de acuerdo con instrucciones específicas.

- **EVALUACIÓN ESTADÍSTICA DE RESULTADOS**

El análisis de los datos recabados con la muestra, incluye indicaciones del valor hasta el cual las estimaciones derivadas de la muestra pueden desviarse de los valores verdaderos de la población. Esta evaluación debe comprender datos sobre la precisión de los estimadores, sobre todo si la selección ha sido probabilística, así como consideraciones sobre algunos sesgos en la operación de reconocimiento que tienda a distorsionar el valor de los estimadores.

Dentro de los sesgos puede considerarse las "no respuestas", cobertura, influencia de los enumeradores sobre la unidad de muestre entrevistado y lo que anoten en el cuestionario, una codificación de respuestas inadecuada, etc.

Por lo que se refiere a la precisión esta se refiere al error de muestreo de un estimador: mientras más pequeño sea el error, mejor será la precisión. El error de muestreo se mide con la fórmula del error estándar, la cual varía de acuerdo con el tipo de estimador - media, mediana, razón, etc. y con el diseño muestral.

La exposición de las fórmulas de los errores estándar se presentan en la sección de los métodos de muestreo, donde se deducen de las varianzas de los estimadores - media, total, etc.

- **DISEÑO DE LOS FORMATOS DE TABULACIÓN :**

Los requerimientos de información y las relaciones significativamente importantes, deben exhibirse en estos formatos con claridad y sencillez, dado que con el éxito que esto se logre, la solución del problema será más convincente y fácil. Deben definirse aquí los títulos de todos los cuadros.

- **DISEÑO DEL CUESTIONARIO E INSTRUCTIVO :**

Basándose en los formatos de tabulación, del guión de información, de los rangos probables de variación, de las experiencias anteriores y de las posibles respuestas de las preguntas, hágase el diseño de un cuestionario precodificado, procurando y verificando que no se omita ningún concepto, que el llenado del cuestionario, sea lo más sencillo y rápido posible, que el encadenamiento de las preguntas sea el más adecuado, que algunas preguntas sirvan para comprobar las respuestas de otras, etc. Un cuestionario precidificado asigna en cada pregunta un conjunto de claves numéricas, correspondiendo en forma biunívoca, en el conjunto de las posibles respuestas, esta claves se anotan cifra por cifra, en las posiciones -en cuadrícula- que se hayan designado para el caso.

- **INVESTIGACIÓN SOBRE FUENTES DE INFORMACIÓN**

Un marco muestral es un conjunto de listas o de mapas, o una combinación de estos elementos, de tal manera, que todas las unidades de interés estén contenidas y que al seleccionar las muestra se pueda determinar la probabilidad de su inclusión, asimismo en el momento de levantar la encuesta, la identificación de cada unidad en la muestra sea posible hacerla sin ninguna ambigüedad.

Para obtener un marco muestral puede recurrirse a ciertas instituciones y recopilar además , datos para: calcular el tamaño de la muestra , confrontar y complementar los resultados de la encuesta, determinar aproximadamente algunos rangos de variación, etc., si es que en los antecedentes -archivos propios- no se tienen.

- **PRUEBA DEL CUESTIONARIO Y AJUSTES FINALES.**

Con objeto de determinar cuáles ajustes deben hacerse al cuestionario para poder lograr los objetivos en forma satisfactoria, es necesario realizar algunas entrevistas en el campo de estudio, llenar los cuestionarios correspondientes y evaluar los resultados a este nivel.

- **FORMULACIÓN DEL GUIÓN DE INFORMACIÓN**

Partiendo de un examen del problema, se recomienda hacer una relación de todas las variables, cuyos valores puedan ser significativamente relevantes, en la resolución del problema.

- **OBTENCIÓN DE INFORMACIÓN COMPLEMENTARIA**

Prepárese todo el material que sea necesario, como oficios debidamente dirigidos y firmados, formas para captar información, etc. Los métodos de muestreo tienen por objeto indicar el número de unidades que deben incluirse en la muestra, dependiendo de la forma como éstas se seleccionen, del nivel de confianza que se requiera, del error de muestreo permisible y del fondo disponible para la realización de la encuesta.

- **LEVANTAMIENTO DE LA ENCUESTA**

El trabajo de los enumeradores debe hacerse exactamente con las unidades de última etapa, determinadas en la selección de la muestra y si ello no fuera posible por deficiencias en el marco muestral, resuélvase el problema con apego a las instrucciones precisas que se hayan hecho para estos pasos. Al hacerse las preguntas, téngase cuidado de que las respuestas sean correctas y veraces, considerando los rangos aproximados para los valores que puedan tomar las variables involucradas en el estudio.

- **SUPERVISIÓN DEL LEVANTAMIENTO DE LA ENCUESTA**

Es conveniente utilizar una forma de reporte, en la cual el supervisor anote cómo se desarrolla el levantamiento de la encuesta, esto es, registrar el material recibido y entregado, folio de los cuestionarios entregados a su grupo, casos de no respuesta y especificación de la resolución tomada, folio de los cuestionarios que fue necesario aclarar, número diario de cuestionarios entregados y de errores por enumerador, porcentaje del avance total del trabajo -llenado de cuestionarios-, día y hora para cada reporte a oficinas centrales, números de cuestionarios efectivamente llenados al terminar la encuesta y registro de los demás documentos recogidos, calificación final de los enumeradores, etc.

- **ADMINISTRACIÓN DEL LEVANTAMIENTO DE LA ENCUESTA:**

Se refiere a todas las actividades, como:

-Autorización de gastos y obtención de fondos junto con las directrices administrativas para su uso;

-Acuse de lo recibido a oficinas centrales;

-Pago del trabajo de campo;

-Observación del sistema de envíos;

-Tiempos transcurridos entre envío y recepción;

-Condición de llegada del material;

-Retroalimentación de las experiencias de la fase inicial y ajuste en donde ello sea necesario;

-Registro de aquéllos procedimientos -o personas- que no funcionaron para referencias futuras y para obtener de ello una experiencia;

-Terminación de obligaciones con el personal eventual; etc.

- **CRÍTICA DE LOS CUESTIONARIOS Y DETERMINACIÓN DEL TAMAÑO EFECTIVO DE LA MUESTRA:**

Esto es, hacer un filtrado de todos los errores que no hayan sido detectados por los supervisores, así como también verificar y concentrar el número total de cuestionarios encomendados a cada supervisor, para obtener el tamaño efectivo de la muestra.

- **ANÁLISIS Y DETERMINACIÓN DE LOS ESTÁNDARES DE TRABAJO:**

Basandose en el trabajo realizado, al probar el cuestionario y en experiencias anteriores, determínese el número de cuestionarios por individuos y por día como: cargo de trabajo, número de visitas antes de declarar la no respuesta, mínimo de los rangos de variación para algunas variables, etc.

- **DETERMINACIÓN DETALLADA DE TODOS LOS CONTROLES ADMINISTRATIVOS Y DE TRABAJO**

Elaboración de todos los mecanismos relativos al procedimiento de pago, métodos de retiro de fondos, de adquisición de materiales, de órdenes de trabajo, autorizaciones necesarias, procedimientos para pago de impuestos, definición de obligaciones por ambas partes, procedimientos de envíos, etc.

Redacción de formas para: para reportes del avance del trabajo, registro de personal - con número de clave del registro federal de causantes -, credenciales de identificación, contratos, etc.

Establecimiento de los requisitos del personal de campo:

escolaridad, disponibilidad de tiempo, sexo, número de ello en las diversas plazas y zonas de la encuesta, preferencia de contratar maestros, trabajadores sociales, enfermeras, etc.

1) ENTRENAMIENTO DE SUPERVISORES:

En todas las actividades, tanto administrativas, como de trabajo de campo.

2) PROCEDIMIENTOS PARA EL VIAJE DE SUPERVISORES:

Habilitación de supervisores de campo con visticos, vehículos, boletos, credenciales, cartas de presentación debidamente dirigidas y firmadas etc.

3) APROVISIONAMIENTO DE MATERIALES PARA TRABAJO:

Como cuestionarios, instructivos, fotografías, planos, cintas métricas, etiquetas, ligas, requisiciones para vehículos, gasolina, papelería, tabletas para escribir en el campo, lápices, gomas, plumones, tarjetas con direcciones importantes y teléfonos, instrucciones para casos de emergencias, formas para control y registro del material, para el avance del trabajo, etc.

4) ORGANIZACIÓN DE LOS MATERIALES DE TRABAJO:

En paquetes para las diversas zonas y unidades; ¿a quién van dirigidos? controles de entradas y salida de documentos.

5) ANÁLISIS ESTADÍSTICO DE LOS RESULTADOS:

Partiendo de una evaluación de la no respuesta y la no cobertura, los intervalos de confianza y los coeficientes de variación se puede obtener un criterio, acerca de la validez de las estimaciones hechas a partir de la muestra, de tal manera que el confrontar algunos resultados de la encuesta, con cifras de otras fuentes, se podrá determinar el origen de las posibles discrepancias significativas, como coberturas distintas, error por no respuesta, error de muestreo, etc.

Este análisis debe de tenerse en cuenta para el reporte final.

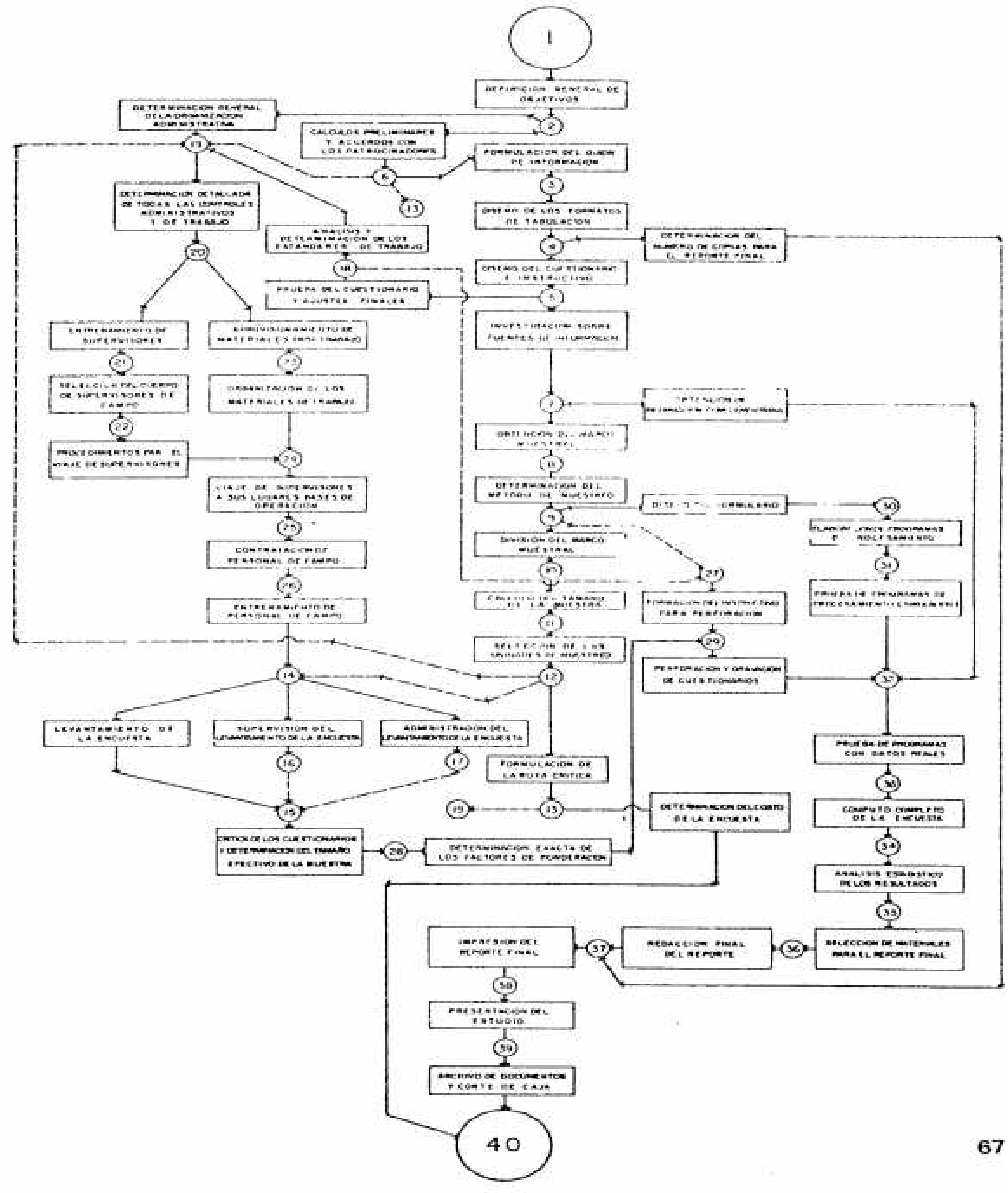
6 ARCHIVO DE DOCUMENTOS

Envío del reporte a la biblioteca para su conservación y clasificación. Destrucción de cuestionarios y otros materiales no utilizables. Control sobre estas operaciones. Felicitaciones al personal involucrado. Cierre de libros y disposición final de fondos. Terminación final de la encuesta.

Por otra parte y considerando que la economía y los negocios son diversos y complejos, decidí incluir para los especialistas una relación adicional de 10 modelos de muestreo, que complementan los anteriores y brindan al lector una gama de alternativas para seleccionar el método apropiado para la investigación especifica que pretenda hacer.

• **RED GENERAL DE ACTIVIDADES EN UNA ENCUESTA DE MUESTREO**

3.1.8 RED GENERAL DE ACTIVIDADES EN UNA ENCUESTA DE MUESTREO



SINTESIS DE LAS CARACTERISTICAS DE ALGUNOS MODELOS DE MUESTREO

DENOMINACION DEL MODELO DE MUESTREO	MEDIA POBLACIONAL	ESTIMADORES: MEDIA POBLACIONAL	ESTIMADORES: VARIANCIA DEL ESTIMADOR DE LA MEDIA POBLACIONAL	VARIANCIA DEL ESTIMADOR DE LA MEDIA POBLACIONAL	TAMAÑO DE LA MUESTRA
Muestreo monoetápico, equiprobable y sin reemplazo (muestreo irrestricto aleatorio).	$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$	$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$Est.\ V(\hat{\mu}) = \frac{N-n}{Nn} s^2$	$V(\hat{\mu}) = \frac{N-n}{Nn} S^2$ $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i-\mu)^2$	$n = \frac{\frac{S^2 t^2}{\varepsilon^2\mu^2}}{1+\frac{1}{N}\frac{S^2 t^2}{\varepsilon^2\mu^2}}$
Muestreo bietápico, equiprobable y sin reemplazo.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Muestreo monoetápico, equiprobable, sin reemplazo y estratificado.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Muestreo bietápico, equiprobable, sin reemplazo y estratificado.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Muestreo monoetápico, equiprobable y con reemplazo (muestreo irrestricto aleatorio con reemplazo).	$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$	$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$	$Est.\ V(\hat{\mu}) = \frac{s^2}{n}$ $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$	$V(\hat{\mu}) = \frac{N-1}{Nn} S^2$	$n = \frac{(N-1)S^2 t^2}{N\varepsilon^2\mu^2}$
Muestreo bietápico, equiprobable y con reemplazo.	[illegible]	[illegible]	$Est.\ V(\hat{\mu}) = \frac{s^2}{nM^2}$ [illegible]	[illegible]	[illegible]
Muestreo monoetápico, equiprobable con reemplazo y estratificado.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Muestreo bietápico, equiprobable, con reemplazo y estratificado.	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Muestreo monoetápico con probabilidades variables de selección y con reemplazo.	$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i$ $\sum_{i=1}^{N} P_i = 1$	$\hat{\mu} = \frac{1}{Nn}\sum_{i=1}^{n}\frac{X_i}{P_i}$	$Est.\ V(\hat{\mu}) = \frac{s^2}{n}$ $s^2 = \frac{1}{N^2(n-1)}\sum_{i=1}^{n}\left(\frac{X_i}{P_i} - N\hat{\mu}\right)^2$	$V(\hat{\mu}) = \frac{1}{n}\sigma^2$ $\sigma^2 = \frac{1}{N^2}\sum_{i=1}^{N} P_i\left(\frac{X_i}{P_i} - N\mu\right)^2$	$n = \frac{\sigma^2 t^2}{\varepsilon^2\mu^2}$
Muestreo bietápico con probabilidades variables de selección en cada etapa con reemplazo.	[illegible]	[illegible]	$Est.\ V(\hat{\mu}) = \frac{s^2}{n}$ [illegible]	[illegible]	[illegible]

- **PRÁCTICA PARA REAFIRMAR EL CONOCIMIENTO**

Una población consta de los dígitos: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, a partir de la cual obtenga:

a) El número y composición de las muestras de tamaño n= 2 que pueda surgir de esa población, con reemplazo y sin reemplazo.

b) Considerando el número de muestras obtenidas sin reemplazo. Obtenga «$\bar{x}$ y ±$_{\bar{x}}$; y para la población obtenga μ y σ.

c) Analice, compare interprete la relación que hay entre μ, «$\bar{x}$, σ, ±$_{\bar{x}}$ tanto para "variables" como para proporciones (atributos).

d) Obtenga $\overline{X}_i$ y *Si* para cada muestra obtenida sin reemplazo.

e) Compare e interprete los valores de los parámetros μ y σ y de los estadísticos $\overline{X}_i$ y Si, si es que existe.

f) Usando la tabla de números aleatorios seleccione aleatoria y sistemáticamente una muestra. Para la muestra seleccionada, calcule su $\bar{x}$, *S*i correspondiente. Con α = 0.05, determine e interprete los limites de confianza dentro de las cuales se halla «$\bar{x}$.

g) Con e = 0.05 (error permitido) y α = 0.05 determine el tamaño de la muestra adecuada, aplicando la fórmula del muestreo simple aleatorio.

h) Relación entre ±$_{\bar{x}}$, e.

<u>SOLUCIÓN DE LA PRACTICA VIII</u>

a) Con X_i: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Tenemos N = 10, luego con n = 2 se obtiene, sin reemplazo

$\begin{bmatrix} N \\ n \end{bmatrix} = \frac{N!}{(N-n)!n!} = \frac{10}{(10-2)!2!} = 45$ muestras de tamaño dos y que constituyen la nueva distribución de muestreo, que son:

Muestra	Composición de la muestra	P(muestra)	Media Muestral	$Si=\sqrt{\frac{\sum\left(xi-\overline{x}\right)^{2}}{n}}$
1	0 , 1	1 ÷ 45	0.5	0.5
2	0 , 2	1 ÷ 45	1.0	1.0
3	0 , 3	1 ÷ 45	1.5	1.5
4	0 , 4	1 ÷ 45	2.0	2.0
5	0 , 5	1 ÷ 45	2.5	2.5
6	0 , 6	1 ÷ 45	3.0	3.0
7	0 , 7	1 ÷ 45	3.5	3.5
8	0 , 8	1 ÷ 45	4.0	4.0
9	0 , 9	1 ÷ 45	4.5	4.5
10	1 , 2	1 ÷ 45	1.5	0.5
11	1 , 3	1 ÷ 45	2.0	1.0
12	1 , 4	1 ÷ 45	2.5	1.5
13	1 , 5	1 ÷ 45	3.0	2.0
14	1 , 6	1 ÷ 45	3.5	2.5
15	1 , 7	1 ÷ 45	4.0	3.0
16	1 , 8	1 ÷ 45	4.5	3.5
17	1 , 9	1 ÷ 45	5.0	4.0
18	2 , 3	1 ÷ 45	2.5	0.5
19	2 , 4	1 ÷ 45	3.0	1.0
20	2 , 5	1 ÷ 45	3.5	1.5

21	2 , 6	1 ÷ 45	4.0	2.0
22	2 , 7	1 ÷ 45	2.5	2.5
23	2 , 8	1 ÷ 45	5.0	.
24	2 , 9	1 ÷ 45	5.5	.
25	3 , 4	1 ÷ 45	3.5	.
26	3 , 5	1 ÷ 45	4.0	.
27	3 , 6	1 ÷ 45	4.5	.
28	3 , 7	1 ÷ 45	5.0	.
29	3 , 8	1 ÷ 45	5.5	.
30	3 , 9	1 ÷ 45	6.0	.
31	4 , 5	1 ÷ 45	4.5	.
32	4 , 6	1 ÷ 45	5.0	.
33	4 , 7	1 ÷ 45	5.5	.
34	4 , 8	1 ÷ 45	6.0	.
35	4 , 9	1 ÷ 45	6.5	.
36	5 , 6	1 ÷ 45	5.5	.
37	5 , 7	1 ÷ 45	6.0	.
38	5 , 8	1 ÷ 45	6.5	.
39	5 , 9	1 ÷ 45	7.0	.
40	6 , 7	1 ÷ 45	6.5	.
41	6 , 8	1 ÷ 45	7.0	.
42	6 , 9	1 ÷ 45	7.5	.
43	7 , 8	1 ÷ 45	7.5	.

44	7 , 9	1 ÷ 45	8.0	.
45	8 , 9	1 ÷ 45	8.5	.
		45 ÷ 45	202.5	

Generación de la distribución de muestras, con reemplazo; $N^n = 10^2 = 100$

n_i	Composición	n_i	Composición	n_i	Composición	n_i	Composición
1	0 , 0	26	2 , 5	51	5 , 0	76	7 , 5
2	0 , 1	27	2 , 6	52	5 , 1	77	7 , 6
3	0 , 2	28	2 , 7	53	5 , 2	78	7 , 7
4	0 , 3	29	2 , 8	54	5 , 3	79	7 , 8
5	0 , 4	30	2 , 9	55	5 , 4	80	7 , 9
6	0 , 5	31	3 , 0	56	5 , 5	81	8 , 0
7	0 , 6	32	3 , 1	57	5 , 6	82	8 , 1
8	0 , 7	33	3 , 2	58	5 , 7	83	8 , 2
9	0 , 8	34	3 , 3	59	5 , 8	84	8 , 3
10	0 , 9	35	3 , 4	60	5 , 9	85	8 , 4
11	1 , 0	36	3 , 5	61	6 , 0	86	8 , 5
12	1 , 1	37	3 , 6	62	6 , 1	87	8 , 6
13	1 , 2	38	3 , 7	63	6 , 2	88	8 , 7
14	1 , 3	39	3 , 8	64	6 , 3	89	8 , 8
15	1 , 4	40	3 , 9	65	6 , 4	90	8 , 9
16	1 , 5	41	4 , 0	66	6 , 5	91	9 , 0
17	1 , 6	42	4 , 1	67	6 , 6	92	9 , 1

18	1 , 7	43	4 , 2	68	6 , 7	93	9 , 2
19	1 , 8	44	4 , 3	69	6 , 8	94	9 , 3
20	1 , 9	45	4 , 4	70	6 , 9	95	9 , 4
21	2 , 0	46	4 , 5	71	7 , 0	96	9 , 5
22	2 , 1	47	4 , 6	72	7 , 1	97	9 , 6
23	2 , 2	48	4 , 7	73	7 , 2	98	9 , 7
24	2 , 3	49	4 , 8	74	7 , 3	99	9 , 8
25	2 , 4	50	4 , 9	75	7 , 4	100	9 , 9

b) Cálculo de los parámetros de la población.

Valores fijos.

μ = 45/10 = 4.5 .

$$\pm = \sqrt{\frac{\sum (x_i - «)^2}{N}} = \sqrt{\frac{82.50}{10}} = \pm = 2.87$$

Calculo de σ:

X_i	X_i-μ	$(X_i-\mu)^2$
0	-4.5	20.25
1	-3.5	12.25
2	-2.5	6.25
3	-1.5	2.25
4	-0.5	0.25
5	0.5	0.25
6	1.5	2.25
7	2.5	6.25

8	3.5	12.25
9	4.5	20.25
	0	82.5

Ahora, calculando.

$$«\bar{x} = E(\bar{X}_i) = \frac{\sum \bar{X}_i}{n} = \frac{202.5}{45} = 4.5 = «\bar{x}$$

NOTA :número de muestras = 45.

$$±\bar{x} = \frac{±}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{2.87}{\sqrt{2}} * \sqrt{\frac{10-2}{10-1}} = \frac{2.87}{1.41}\sqrt{0.89}$$

$$±\bar{x} = 2.03(0.94) = 1.92$$

c) Vemos que «$\bar{x}$ = μ = 4.5 mientras que $±_{\bar{x}}$ = 1.92 y que σ = 2.87

En el caso de "VARIABLES", lo mismo sucede en el caso de las "PROPORCIONES", es decir, P = Π donde :

P: Proporción de la distribución de muestras.

Π: Parámetro poblacional.

Sp: Error estándar de la distribución de proporciones.

σp: Parámetro denominado desviación estándar de la población.

Sp diferente de σp.

d) Obtenga $\bar{X}_i$ y si para cada muestra obtenida <u>sin reemplazo</u>.

Solución:

i = 1, 2,...,44, 45.

Calculando como ilustración S1 y S45, porque el método de calculo es el mismo, tenemos:

$$S_1 = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$$
$$S_1 = \sqrt{\frac{\sum \left[(0-0.5)^2 + (1-0.5)^2\right]}{2}}$$
$$S_1 = \sqrt{\frac{\sum [0.25 + 0.25]}{2}}$$
$$S_1 = \sqrt{\frac{0.50}{2}}$$
$$S_1 = \sqrt{0.25} = 0.5$$

$$S_{45} = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}}$$
$$S_{45} = \sqrt{\frac{\sum \left[(8-8.5)^2 + (9-8.5)^2\right]}{2}}$$
$$S_{45} = \sqrt{\frac{\sum [0.25 + 0.25]}{2}}$$
$$S_{45} = \sqrt{\frac{0.50}{2}}$$
$$S_{45} = \sqrt{0.25} = 0.5$$

e) Al comparar los valores de μ, σ con $\overline{X}_i$, Si, vemos que el valor de los parámetros es FIJO, mientras que el de las "estadísticas" es variable puesto que esta en función de la composición de cada muestra.

f) La selección aleatoria determinó la obtención de la muestra compuesta por los dígitos 0 y 8, puesto que la tabla de números aleatorios, trabajando horizontalmente, determinó que se tomara la muestra número 08 del marco muestral que está compuesto por 45 muestras disponibles y obtenidas en un muestreo sin remplazo.

MARCO MUESTRAL	
Número de muestra	Composición de la muestra
1	0 , 1
2	0 , 2
3	0 , 3
4	0 , 4
5	0 , 5

6	0 , 6
7	0 , 7
8	0 , 8
.	.
.	.
.	.
.	.
.	.
45	8 , 9

Así, a partir de la selección aleatoria que determinó la muestra compuesta por los dígitos 8 y 0 vamos a determinar los límites de confianza con: = 0.05; Z = ± 1.96.

$\bar{X}$ = 8 + 0 / 2 = 4

S = $\sqrt{\frac{\sum\left[(8-4)^2+(0-4)^2\right]}{2}} = \sqrt{16} = 4$ = 4

Sabemos que los límites de confianza se determinan con :

$\bar{X} \pm Z_{\sim} \pm \bar{x}$ donde $\pm\bar{x} = \frac{\pm}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = 1.92$ del inciso B).

Luego sustituyendo tendremos:

4 - (1.96) (1.92) límite inferior del intervalo de confianza = 0.2368

4 + (1.96)(1.92) límite superior del intervalo de confianza = 7.7632.

Interpretación: hay una probabilidad del 95% de que el valor de se halle en el intervalo de 0.2368 a 7.7632, lo cual es cierto puesto que «$\bar{x}$ = 4.5.

g) Con e = 0.05; α = 0.05 tenemos Zα = ± 1.96; σ = 2.87 luego:

$$n = \frac{Z_{\sim}^2 \pm_x^2 N}{e^2 N - e^2 + Z_{\sim}^2 \pm_x^2} = \frac{(3.84)(3.68)(10)}{(0.25) - (0.0025) + (14.15)} = \frac{141.7}{14.17} = 10$$

Ahora bien usando $n = \dfrac{N}{1 + Ne^2} = \dfrac{10}{1 + 10(0.05)^2} = \dfrac{10}{1 + 0.025}$

$$n = \frac{10}{1.025} = 9.7 \cong 10$$

Observaciones:

Con las dos fórmulas se obtiene el mismos resultado. Ello indica que el tamaño de la muestra debe ser el del universo. Esto es así, no debe sorprendernos porque el universo es tan pequeño que la muestra debe ser igual a 10 para que sea representativa.

h) La relación entre $\pm_{\bar{x}}$ y e.

$$\pm\bar{x} = \frac{\pm}{\sqrt{n}} * \sqrt{\frac{N-n}{N-1}} = \frac{2.87}{\sqrt{2}} * \sqrt{\frac{10-2}{10-1}} = 1.92$$

e = $Z_{\sim}\pm\bar{x}$ = 1.96 (1.92) = 3.7632

i).- La determinación de e se hace a partir de los límites de confianza.

Así, de $\bar{X} \pm Z_{\sim}\pm\bar{x}$ tenemos que e = $Z_{\sim}\pm\bar{x}$ = 1.96 (1.92) = 3.7632

Comparación $\pm_{\bar{x}} = 1.92$ y e = 3.7632, luego el error estándar, es menor que el error de muestreo o error permitido.

Pero si el error estándar (1.92) lo usamos en términos de probabilidad para cuantificar el error de muestreo $|\bar{x} - «_{x}|$, entonces recordemos que **idealmente** éste último debe ser menor o igual que e = error permitido= $Z_{\sim} \pm_{\bar{x}}$

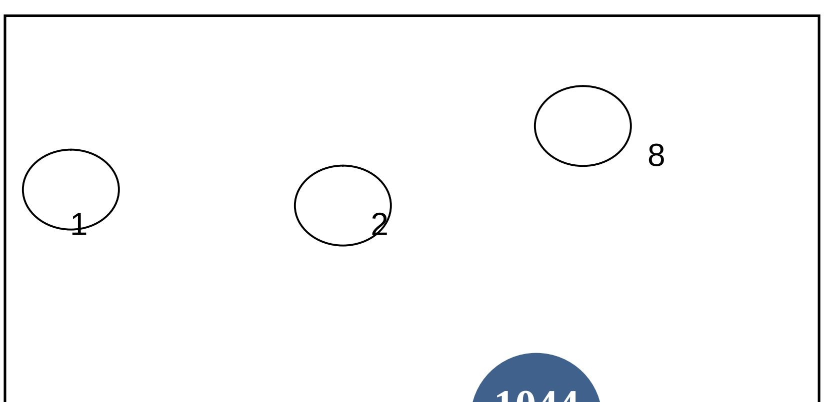
Del inciso b), sabemos que $«_{x} = 4.5$ y del inciso f) sabemos que $\bar{x} = 4.0$ luego el error de muestreo = $|4 - 4.5| = |0.5| \leq 3.7637$ = error permitido. Es bueno el resultado.

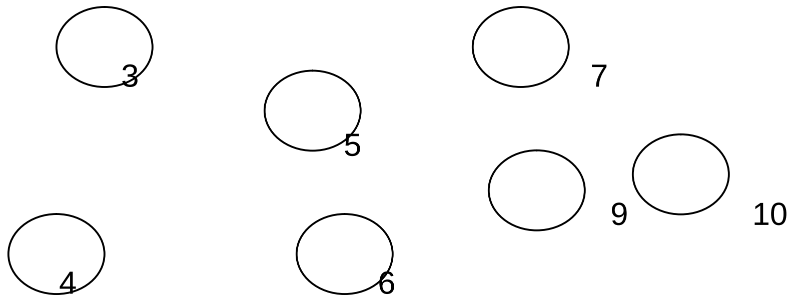
- **PRÁCTICA DOS**

1.- APLICACIONES DEL MUESTREO SIMPLE ALEATORIO.

Referencias:

El plano de la colonia del Valle del Distrito Federal es el siguiente:





Con : n = 2

t = 2

Obtenga:

a) El número de manzanas en la colonia del Valle.

hay N = 10 manzanas.

b) La fracción del muestreo.

Fracción de muestreo = F = n / N = 2 / 10 = 0.2

c) Seleccione con la tabla de números aleatorios las dos manzanas

que integren la muestra, indica como son y como le hizo.

Digamos que fueron las manzanas 2 y 7, que cayeron en la muestra mediante el manejo ya conocido de la tabla de números aleatorios.

d) Suponiendo que:

La primer manzana tiene 40 familias.

La segunda manzana tiene 36 familias.

Cálcule la media y desviación estándar de la muestra.

Puesto que:

Manzana 2 tiene 40 familias.

Manzana 7 tiene 36 familias.

–

X = 40 + 36 / 2 = 76 /2 = 38 familias.

$$S = \sqrt{\frac{(40-38)^2 + (36-38)^2}{2}} = \sqrt{\frac{(2)^2 + (-2)^2}{2}} = \sqrt{\frac{8}{2}} = 2$$

e) Determine el total de familias en la colonia del Valle.

El total de familias se estima por :

$\hat{Y} = N\overline{X}$; $\hat{Y} = 10(38);$

$\hat{Y}$ = 380 familias.

f) Determine e interprete los límites de confianza del total de familias.

Sabemos que:

$$N\overline{x} - \frac{tNS}{\sqrt{n}}\sqrt{1-F} \leq \hat{Y} \leq N\overline{x} + \frac{tNS}{\sqrt{n}}\sqrt{1-0.2}$$

Sustituyendo

$$380 - \frac{2(10)2}{1.41}\sqrt{1-0.2} \leq \hat{Y} \leq 380 + \frac{2(10)2}{1.41}\sqrt{1-0.2}$$

$$380 - \frac{2(10)2}{1.41}(0.89) \leq \hat{Y} \leq 380 + \frac{2(10)2}{1.41}(0.89)$$

380 - 28.37 (0.89) | $\hat{Y}$ | 380 + 28.37 (0.89)

380 - 25.2 | $\hat{Y}$ |380 + 25.2

354.8 | $\hat{Y}$ |405.2

Interpretación:

El total estimado de familias Y, se halla entre 355 y 405 familias

con una probabilidad o seguridad del 95.45%

g) Determine el número de habitantes en la colonia del Valle tomando en cuenta que 5 es el promedio de personas por familia.

El total de habitantes en la colonia del Valle es:

Total de habitantes = 380 (5) = 1900 personas.

- **APLICACIÓN DEL MUESTREO ALEATORIO ESTRATIFICADO:**

Referencias: El canal 22 de televisión ha sido puesto en venta y la empresa "escorpión" que esta interesado en adquirirlo disidió hacer una encuesta para conocer el número de horas que el público ve televisión y de ahí saber cuántos hogares (mediante entrevistas) ven el canal 22.

La empresa escorpión puede producir estimaciones por separado es decir, puede estratificar para estimar el número promedio de horas que se ve televisión en cada estrato, ya que, la información disponible revela que hay tres estratos que componen el universo o población con:

POBLACIÓN N	MUESTRA n	ESTRATO HOGARES
N_1= 180 hogares N_2= 70 hogares N_3= 100 hogares	n_1= 15 hogares n_2= 4 hogares n_3= 5 hogares	1 2 3
Total 350	Total 24	

Mediante la cual se realizan las entrevistas, con los siguientes resultados:

Tiempo que se ve televisión, en horas por semana.

Estrato 1	Estrato 2	Estrato 3
30, 27, 40 45, 26, 35 33, 29, 37 34, 25, 41 43, 32, 31	4, 49 25, 30	9, 20 11, 34 24,

Con esos datos sustituya y obtenga:

a) $\bar{X}_i$, Si con i=1, 2, 3.

Estrato 1	Estrato 2	Estrato 3
n_1 = 15	n_2 = 4	n_3 = 5

$\overline{x}_1 = 34$	$\overline{x}_2 = 27$	$\overline{x}_3 = 20$
$S_1 = 6$	$S_2 = 16$	$S_3 = 9.1$
$N_1 = 180$	$N_2 = 70$	$N_3 = 100$

ESTRATO 1

$$\overline{X}_1 = \frac{30+27+40+45+26+35+33+29+37+34+25+41+43+32+31}{15}$$

$$\overline{X}_1 = \frac{508}{15} = 33.87$$

33.87 equivalente a 34

$$S_1 = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}} = \sqrt{\frac{546}{15}} = \sqrt{36.4} = 6$$

$\overline{X}_i$	$x_i - \overline{X}_1$					$(X_i - \overline{X}_1)^2$
30	30	-	34	=	-4	16
27	27	-	34	=	-7	49
40	40	-	34	=	6	36
45	45	-	34	=	11	121
26	26	-	34	=	-8	64
35	35	-	34	=	1	1
33	33	-	34	=	-1	1

29	29	-	34	=	-5	25
37	37	-	34	=	3	9
34	34	-	34	=	0	0
25	25	-	34	=	-9	81
41	41	-	34	=	7	49
43	43	-	34	=	9	81
32	32	-	34	=	-2	4
31	31	-	34	=	-3	9
						546

ESTRATO 2

$$\overline{X}_2 = \frac{4+49+25+30}{4} = \frac{108}{4} = 27$$

$$S_2 = \sqrt{\frac{\sum (x_i - \overline{x})^2}{n}} = \sqrt{\frac{1026}{4}} = \sqrt{256.5} = 16.0$$

$\overline{X}_i$	$X_i - \overline{X}_1$					$(X_i - \overline{X}_2)^2$
4	4	-	27	=	-23	529
49	49	-	27	=	-22	484
25	25	-	27	=	-2	4
30	30	-	27	=	3	9

1026

ESTRATO 3

$$\overline{X}_3 = \frac{9+20+11+34+24}{5} = \frac{98}{5} = 19.6 \text{ equivale a } 20$$

$$S_3 = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} = \sqrt{\frac{414}{5}} = \sqrt{82.8} = 9.1$$

$\bar{X}_1$	$X_1 - \bar{X}_3$					$(X_1 - \bar{X}_3)^2$
9	9	-	20	=	-11	121
20	20	-	20	=	0	0
11	11	-	20	=	-9	81
34	34	-	20	=	14	196
24	24	-	20	=	4	16
						414

b) Usando la información anterior, estime el tiempo promedio que se ve televisión en horas por semana, para todos los hogares, en la población constituida por todos, sabiendo que esta medida poblacional, que no es muestral, se calcula con la fórmula:

$\bar{\bar{X}}$ De todos los estratos $= \frac{1}{N}\sum_{i=1}^{3} N_i \bar{X}_i = \frac{1}{N}\left[N_1\bar{X}_1 + N_2\bar{X}_2 + N_3\bar{X}_3\right]$

$$\bar{\bar{X}} = \frac{1}{350}\left[180(34) + 70(27) + 100(20)\right]; \bar{\bar{X}} = \frac{1}{350}\left[6120 + 1890 + 2000\right]$$

$\bar{\bar{X}} = \frac{10{,}010}{350} = 28.6$ equivalente a 29 horas

c) Obtenga la varianza de $\bar{\bar{X}}$ con la fórmula:

$$V(\bar{\bar{X}}) = \frac{1}{N^2}\left[N_1\left(\frac{N_1 - n_1}{N_1}\right)\left(\frac{S^2}{n_1}\right) + N_2\left(\frac{N_2 - n_2}{N_2}\right)\left(\frac{S_2^2}{n_2}\right) + N_3\left(\frac{N_3 - n_3}{N_3}\right)\left(\frac{S_3^3}{n_3}\right)\right]$$

$$V(\bar{\bar{X}}) = \frac{1}{(350)^2}\left[180\left(\frac{180-15}{180}\right)\left(\frac{36.4}{15}\right) + 70\left(\frac{70-4}{70}\right)\left(\frac{256.5}{4}\right) + 100\left(\frac{100-5}{100}\right)\left(\frac{82.8}{5}\right)\right]$$

$$V(\bar{\bar{X}}) = \frac{1}{122500}\left[402.41 + 4219.75 + 15.732\right]$$

$$V(\bar{\bar{X}}) = \frac{6195.36}{122500} = 0.0506$$

d) Obtenga los limites de confianza con:

$\bar{\bar{X}} \pm 2\sqrt{V(\bar{\bar{X}})}$;   Interprete conociendo que.

$2\sqrt{V(\bar{\bar{X}})}$ = e = error de estimación permitido

$\bar{\bar{X}} \pm 2\sqrt{V(\bar{\bar{X}})}$ $29 \pm 2\sqrt{0.0506}$

29 ± 2 (0.22) ;   29 ± 0.44; luego e = 0.44

Limite inferior = 28.56

Limite superior = 29.44

Interpretación:

Usando el muestreo aleatorio estratificado hemos estimado que el número promedio de horas por semana que se ve televisión en todos los hogares es de 29 horas, el error de estimación permitido es de 0.44 horas, con una probabilidad de 95.45% .

Calificación:

Solución del caso No.1 ; 34 puntos

Solución del caso No.2 ; 66 puntos

total: 100 puntos

EXAMEN SOBRE MUESTREO PARA REAFIRMAR SUS CONCEPTOS Y APLICACIONES.

Nombre del alumno(a)______________________________________Calif___

1.- ¿Con qué fórmula se obtiene el tamaño de la muestra probabilística cuando no se conoce o dispone de ningún dato estadístico?

Respuesta: $n = \frac{1}{e^2}$, donde e es el error permitido, al cuadrado.

2.- Al trabajar con estadística de atributos, usando el método del muestreo simple aleatorio, sabemos que la fórmula del *error permitido* para p=0.5 es

$e = \sqrt{\frac{pq}{n} * Z\sim}$, dicho en otra forma: $e = \pm p * Z\sim$

Con esa referencia calcúlelo con α=10% y n= 400

Respuesta:

Primero usted debe de observar que Z_α está dentro del radical de la raíz cuadrada, lo cual debe causarle extrañeza porque no es congruente con la otra fórmula, por lo que debe hacerlo notar e indicar que ante la duda sacará Z_α del radical. Así, los cálculos son:

$$e = \sqrt{\frac{(0.5)(0.5)}{400}} * 1.65$$
$$e = 0.025(1.65) = 0.04125$$

3.- Si en la estadística de variables el error permitido e= = Zα$\sigma \overline{x}$ y el error de muestreo es $|\overline{x} - \text{«}x|$, tal que $\pm\overline{x} = \frac{\overline{x} - \text{«}\overline{x}}{\pm\overline{x}}$, demuestre que son iguales.

Respuesta: e = Zα $\pm\overline{x} = \frac{\overline{x} - \text{«}\overline{x}}{\pm\overline{x}} * \frac{\pm x}{\sqrt{n}} = \frac{\overline{x} - \text{«}\overline{x}}{\pm\overline{x}} \pm\overline{x} = \overline{x} - \text{«}x$

4.- Derivado de la pregunta 3, determine e interprete el intervalo de confianza dentro del que se halla $\text{«}x$: consumo promedio de leche diario por familia en el Distrito Federal, sabiendo que INEGI con una muestra simple aleatoria de 36 familias encontró que cada una de ellas consume 2 litros de leche/día en promedio y que por estudios previos la σ= 0.2 litros/día. Así, con α= 0.045 y por ello Zα$= \pm 2$,

BIBLIOGRAFÍA.

1. Box, George E.P. and Jenkins, Gwilyn, ***Time Series Analysis: Forecasting and Control***, www.infobolsa.es, 1976

2. Cabrera, Sancho y Serrano. ***Microeconometría y decisión***. Ediciones Pirámide. Madrid, España. 2001. pp. 46-48.

3. Carrascal, Arranz Ursicino, González, González Yolanda, Rodríguez y Prado Beatriz, ***Análisis Econométrico con E-Views***, México, Alfaomega Ra-Ma, 2001.
4. Coro, Chasco Yrigoyen. Departamento de Economía aplicada – Universidad Autónoma Metropolitana (UAM).

5. Del Mar, María, Zamora, mariam.zamorauah.es, 2002.

6. Fregoso, Iglesias Margarita, ***Diplomado La Educación en el Paradigma del Aprendizaje***. Facultad de la Economía de la UNAM, 2008.

7. Granger, Clive W. J. ***Análisis de Series Temporales, Cointegración y Aplicaciones***, Revista Asturiana de Economía, RAE No 30 2004.

8. Granger, Clive W.J. and Newfold, Paul, ***Forecasting Economic TimeSeries,*** Academic Press, March, 1977, Second Edition 1986.

9. Guisán, M. Carmen, ***Causalidad y Cointegración en Modelos Econometricos***, U.S.C., eccgs@usc.es, 2002.

10. Gujarati, Damodar N, ***Econometría,*** Quinta Edición, 2010, ***Econometría***, México, McGraw Hill.

11. Gujarati, Damodar N. ***Econometría***, cuarta edición. McGraw Hill, 4ª edición, julio 2004.

12. Haeussler, Ernest F. ***Matemáticas para Administración y Economía***, México, Pearson Education, 2003, p. 28, 31, 35, 37 y 121

13. Hernández Alonso, José. ***Ejercicios de Econometría***, Madrid, ESIC Editorial, 1992, p. 332.

14. Hernández, José y Herrador, María del Mar. ***Econometría de series temporales***. Ed. Universitas. Madrid, España. 2007.

15. İnsel, Anyul, ***Econometrics time series analysis and econometric modeling***, 2007.

16. Johnston, J &Dinardo, J. ***Métodos de econometría***, Vicens Vives, 2001

17. Lind, Douglas A., Marchal, William G, Wathen, Samuel A. **Estadística Aplicada a los Negocios y a la Economía**, 12o Edición. McGraw Hill, 2005.

18. Loria, Eduardo, **Econometría con aplicaciones,** Prentice Hall, 2003

19. Maddala, G. S. ***Introducción a la Econometría***, México, Prentice Hall Hispanoamericana, 1996, Cap. 1.

20. Mason, Robert, Lind, Douglas A., Marchal, William G., ***Estadística para Administración y Economía***, México, Alfaomega Editor, 10º Edición, 2001, p. 795.

21. Montero, Granados Roberto, ***Variables no estacionarias y cointegración, Universidad de Granada***, www.ugr.es, 2007.

22. Peña, Daniel. ***Análisis de series temporales***, España, Editorial Alianza, 2005, p. 592.

23. Pérez, Julián, ***1.1. Caracterización de un Modelo Económico*****,** Instituto Tecnológico de Estudios Superiores de Monterrey, 1998.

24. Pérez, López César. ***Econometría Avanzada. Técnicas y Herramientas***, España, Pearson Prentice Hall, 2008,

25. Pérez, López César. ***Econometría básica técnicas y herramientas***, España, Pearson Prentice Hall, 2007.

26. Pérez, López César. ***Econometría de las series de tiempo*****,** España, Pearson Prentice Hall, 2006.

27. Pulido, Antonio. ***Predicción y simulación aplicada a la economía y a la gestión de empresas***, España, Pirámide, 1999.

28. Pulido, San Román Antonio, Pérez, García Julián. ***Modelos Econométricos: Guía para la elaboración de modelos econométricos con Eviews.*** España, Ediciones Piramide, 1999.

29. Reyes, De la Rosa José Alberto. ***Apuntes sobre modelos ARIMA***, FE-UNAM. 2009.

30. Richmond, Samuel B. ***Statistical Analysis***, The Ronald Press Company, New York, capítulos de probabilidad, inferencia estadística, regresión y correlación 1964.

31. Rojas, Escudero Joel, ***Curso-Taller VI de Econometría***, Facultad de Economía, UNAM, 2009.

32. Salvatore, Dominick. ***Econometría***, México, Mc Graw Hill, 1993, Cap. 6-8.

33. Sánchez Barajas, Genaro. ***Introducción a la Econometría***, México: UNAM, Facultad de Economía, 2012.

34. Sánchez Barajas, Genaro. ***La Estadística como Método de Análisis Económico***, un enfoque de la sociedad del conocimiento, Grupo Editorial Sagitario, A.C., 2011.

35. Shao, Stephen P. ***Estadística para Economistas y Administradores de Empresas***, México, Herrero Hermanos, 1975, Cap. 21-23.

36. Suriñach, Jordi Caralt. ***Análisis Económico Regional, nociones básicas de la teoría de la cointegración***, España, Antoni Bosch Editor, 1995, p.185

37. Tapia Tovar Gabriel –Carmona Rocha José Gerardo. ***Modelización regional: técnica aplicada al desarrollo regional, en Desarrollo Regional, Local y Empresas***, ININEE,UMSNH, PP. 66 Y 67, 2004.

38. Tello, Carlos. ***Estados y desarrollo económico: México 1920-2006***, México: UNAM, Facultad de Economía, 2007, p. 776

39. Universidad Autónoma de Madrid, **Curso de Predicción Económica y Empresarial**,www.uam.es/predysim,Edición 2004.

40. Users Guide Eviews 4.0**, *Chapter 13: Time Series Regression,*** pp.303-364.

41. Wikipedia, 2003.

42. Wooldridge, Jeffrey M. ***Introducción a la Econometría, un enfoque moderno***, México, Cengage Learning, 2010.

43. Yahoo Finanzas (www.yahoo.com).