

Universidad de Los Andes
Facultad de Ciencias Económicas y Sociales
Escuela de Economía
Cátedra de Métodos Cuantitativos
Econometría I

INDICE GENERAL

- 1. Aspectos Teóricos Básicos**
- 2. Introducción al Uso del Eviews**
- 3. Etapa I: Especificación del Modelo**
- 4. Etapa II: Recolección de Datos**
- 5. Etapa III: Estimación del Modelo**
- 6. Etapa IV: Evaluación del Modelo**
 - 6.1 Evaluación Económica**
 - 6.2 Evaluación Estadística**
 - 6.3 Evaluación Econométrica**
- 7. Etapa V: Capacidad Predictiva del Modelo**

TÓPICOS DE INTERES

- 1. Variables Cualitativas**
- 2. Análisis de Cointegración**

1. Aspectos Teóricos Básicos

1.1. Econometría

Es una herramienta de la Economía que se encarga de medir o cuantificar los hechos económicos a través del uso de la matemática y la estadística; con el propósito de evaluar y/o formular políticas económicas y posteriormente realizar predicciones. Con la econometría, los economistas cuentan con una herramienta basada en conceptos estadísticos y matemáticos que les permite verificar la teoría económica y sus principios (Muñoz, 2000).

1.2. Propósitos de la Econometría

La econometría tiene como objetivo central evaluar las principales hipótesis de la teoría económica, haciendo uso de la estadística y de la matemática, con el fin de:

1ro. Analizar el comportamiento de las principales variables económicas, a través de la cuantificación de ese comportamiento.

2do. Evaluar políticas económicas, para así mantenerlas o cambiarlas de acuerdo a las necesidades económicas y del comportamiento que demuestran las variables de estudio.

3ro. Predecir el comportamiento futuro de esas variables, basándose en el comportamiento pasado, bajo determinadas condiciones.

1.3. Etapas del Proceso Econométrico

La econometría es una técnica que se lleva a cabo como un proceso, y como todo proceso está compuesto por etapas, a saber:

1ro. Especificación del modelo.

2do. Recolección de los datos.

3ro. Estimación del modelo.

4to. Evaluación del modelo.

5to. Predicción y/o evaluación de políticas.

2. Introducción al Uso de Eviews 5

2.1. Iniciando el Eviews

El Eviews se puede iniciar desde el menú inicio, un acceso directo en el escritorio o cualquier otro acceso que se disponga en la computadora. Aparecerá entonces, como se muestra en la figura 1, la ventana principal del programa.

Figura 1

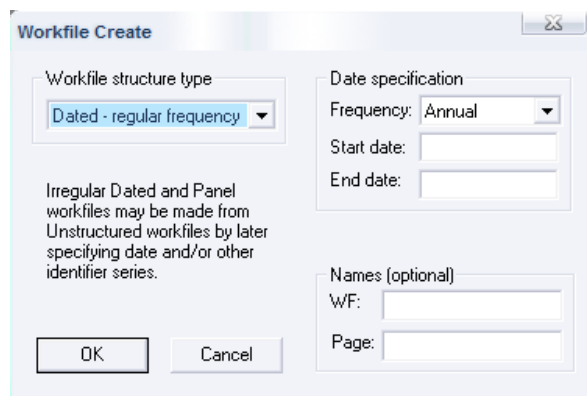


Los elementos que se distinguen son, la **Barra de Título**, que contiene en su parte izquierda el nombre del programa y, en la parte derecha tres botones: minimización, restauración y cierre del Eviews, nombrados de izquierda a derecha según se ven en la figura 1. El **Menú Principal**, que contiene los diferentes comandos (*File, Edit, Object, View, Proc, Quick, Options, Window y Help*) con los cuales se puede operar el programa. La **Ventana de Comandos**, desde la cual se pueden ejecutar la mayoría de los comandos del Eviews. El **Área de Trabajo**, en la cual aparecerán ventanas con los resultados como consecuencia de la ejecución de los comandos. Y, finalmente, la **Línea de Estado**, que de izquierda a derecha, en la primera parte ocasionalmente contiene mensajes para el usuario, en la segunda parte indica la dirección en la cual el programa por defecto busca o graba archivos, en la tercera parte el nombre de la base de datos y, en el extremo derecho el nombre del archivo de trabajo actual.

2.2. Iniciando un workfile (WF) o archivo de trabajo

Una vez iniciado el Eviews, la ruta a seguir desde el menú principal es: **File/New/Workfile...** Al ejecutar el mencionado comando, emergerá un cuadro de diálogo como el que se observa en la figura 2, donde se debe especificar la estructura del archivo de trabajo y, el tipo y frecuencia de los datos.

Figura 2



En la parte izquierda de la figura 2, se distingue la frase “*Workfile structure type*”, con una pestaña de menú desplegable debajo de la misma y, en el cual se debe seleccionar la estructura del archivo de trabajo. Las opciones disponibles son:

1ro. “*Date-regular frequency*”, para datos de series de tiempo, caso en el que las características de la ventana son como las que en efecto se muestran en la figura 2. En la sección “*Date specification*”, se debe seleccionar la frecuencia de la serie de tiempo desde el menú desplegable que emerge en la opción “*Frequency*”; e introducir el primer y último dato en las opciones “*Start date*” y “*End date*”, respectivamente, lo cual se hace de acuerdo a la frecuencia de los datos como se muestra en el cuadro a continuación:

Frecuencia	Descripción	Start Date (Primer Dato)	End Date (Último Dato)	Observaciones
Annual	Anual	1970	2005	Años continuos unos tras otros.
Semi-annual	Semestral	1970/1	2005/2	El número seguido de la barra indica el semestre.
Quarterly	Trimestral	1970/1	2005/4	El número seguido de la barra indica el trimestre.
Monthly	Mensual	1970/1	2005/12	El número seguido de la barra indica el mes.
Weekly	Semanal	1/5/1970	12/30/2005	El primer número es el mes, el siguiente el día en que comienza la semana, el último el año.
Daily [5]	Diario de 5 días (no incluye fines de semana)	1/5/1970	12/30/2005	Se diferencia del anterior al momento de la introducción de los datos, en la tabla para la introducción el número de la observación.
Daily [7]	Diario de 7 días (incluye fines de semana)	1/5/1970	12/30/2005	El primer número es el mes, el segundo el día y el último es el año.

2do. “Unstructured/Undated”, para datos corte transversal, caso en el que la sección “Date specification” se modifica a “Date range”, como se puede detallar en la figura 3, en la cual, en vista de que los datos son atemporales solo basta colocar el número de observaciones que contiene la muestra de la base de datos a utilizar.

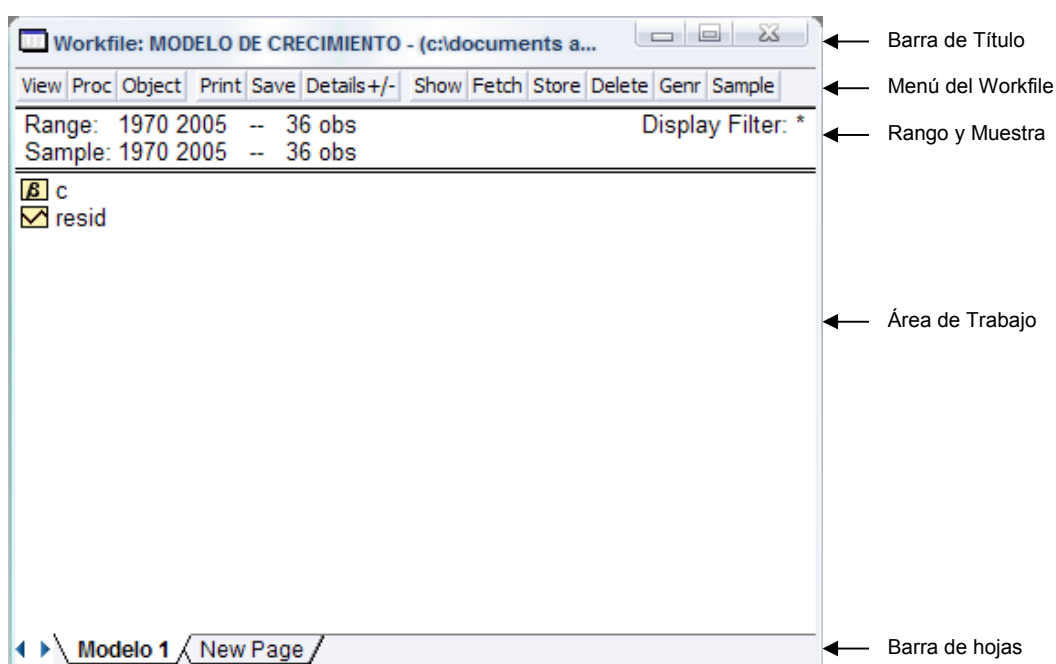
Figura 3

3ro. “Balanced Panel”, para datos en panel, no se utilizan en Econometría I.

Las opciones de la sección “Names” no son necesarias colocarlas, sin embargo “WF” es un espacio dispuesto para colocar el nombre del archivo de trabajo a crear y “Page” para el nombre de la página de ese archivo de trabajo, aunque, esto no significa que el documento este guardado. Para saber como guardarlo diríjase al apartado de este manual desarrollado a este procedimiento.

Al hacer clic en “ok”, se creará el archivo de trabajo o workfile, como el que se puede observar en la figura 4.

Figura 4



Es de hacer notar, que con la creación del archivo de trabajo, aparece un nuevo menú al alcance del usuario para operar el programa, a saber, **Menú del Workfile**. Más abajo se muestra el rango y la muestra de la base de datos utilizada y el área de trabajo. El Eviews se reserva la variable “resid” donde se guardarán los residuos de la última estimación realizada y la variable “c” que funcionará como intercepto para cálculos propios del programa.

2.3. Archivos de trabajos multi-páginas

Una de las bondades del Eviews 5, es que se pueden crear varios archivos de trabajos, cada uno con sus características particulares en cuanto a tipos de datos y tamaño de muestra.

En la opción "Specify by Frequency/Range..." del menú desplegable que emerge al hacer clic con el botón secundario sobre la pestaña "New Page", se da paso a la creación del nuevo archivo de trabajo. Luego, se debe repetir el procedimiento expuesto en "iniciando un workfile".

Si se desea modificar el nombre de alguna de las páginas del workfile, se debe hacer clic con el botón secundario sobre el nombre de la página que desea modificar y saldrá un menú desplegable contentivo con la opción de renombrar (Rename Workfile Page...) y en la ventana emergente introdúzcase el nombre por el cual se desea sustituirlo. Otras opciones del menú desplegable son: borrar la página (Delete Workfile Page) o guardar la página (Save Workfile Page...) del archivo de trabajo.

2.4. Importación de datos

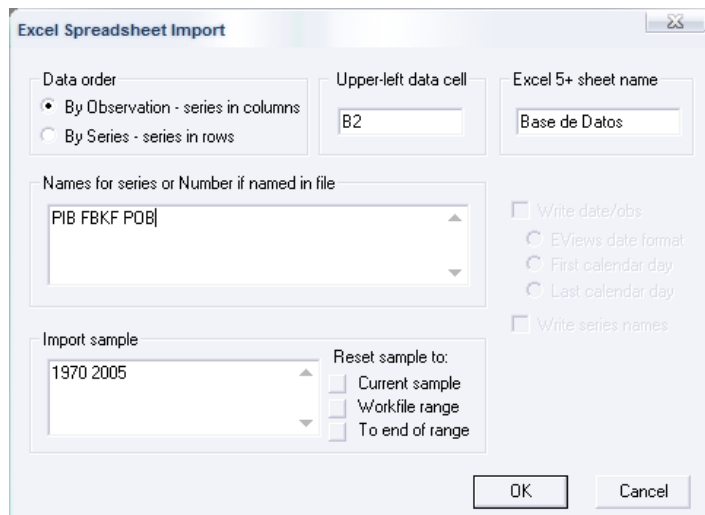
Para importar datos desde una hoja de cálculo del formato Excel hasta el Eviews, es necesario que el documento Excel que contiene la base de datos este guardado y cerrado; es por ello que vale la pena precisar algunos elementos con papel y lápiz a la mano para poder realizar la importación; estos elementos son:

- 1ro. La ubicación en el computador de donde esta guardado el documento.
- 2do. Si los datos están dispuestos en columnas o en filas.
- 3ro. El nombre de la celda donde comienza el primer dato.
- 4to. El nombre de la hoja de cálculo que contiene los datos.
- 5to. En que orden están dispuestas las variables de izquierda a derecha para datos organizados en columnas y de arriba a bajo para datos organizados en filas.
- 6to. El tamaño de la muestra a importar.

Una vez conocidos estos elementos, la ruta a seguir desde el menú principal es: **File/Import/Read Text-Lotus-Excel**. Luego de ejecutado este comando, emergerá una ventana, en donde se debe indicar la ubicación en el computador donde esta guardado el documento Excel contentivo de la base de datos.

Al hacer clic en abrir, aparecerá un cuadro de diálogo como el que se muestra a continuación:

Figura 5



En la sección “Data order” se indica si los datos están organizados en columnas (By Observation – series in columns) o en filas (By Series – series in rows); en el apartado “Upper-left data cell” se introduce el nombre de la celda donde comienza el primer dato de la base de datos; en “Excel 5+ sheet name” se coloca el nombre de la hoja de cálculo contentiva de la base de datos; en la sección “Names for series or Number if named in file” se escriben los nombres de las variables según el orden en que estén dispuestas en la base de datos separadas por un espacio cada una, como se puede ver en la figura 7 y, finalmente en el apartado “Import sample” se coloca el tamaño de la muestra a importar, para luego hacer clic en “OK”.

En el archivo de trabajo aparecerán los nombres de las variables y haciendo doble clic sobre ellas se pueden observar los datos importados.

2.5. Abrir las series como grupo

Con el botón CTRL presionado, se hace clic sobre los nombres de las variables en el orden que se desea aparezcan, luego se deja de presionar CTRL, para hacer clic con el botón secundario sobre algunas de las variables y, en el menú desplegable que aparece seleccionar “Open/as Group”.

Se abrirá una nueva ventana con su propio menú, a saber, **Menú de Trabajo**, el número/fecha (obs) que corresponde a cada observación y, los nombres de las variables con sus respectivos datos (véase figura 6).

Figura 6

obs	PIB	FBKF	POB
1970	20919792	6299435.	10721.00
1971	21562251	7066943.	11123.00
1972	22264618	8124999.	11540.00
1973	23657371	8887905.	11973.00
1974	25091852	8668123.	12422.00
1975	26614335	10890675	12734.00
1976	28948672	13991776	13205.00
1977	30894642	18043344	13693.00
1978	31555280	18761948	14199.00
1979	31976700	15001503	14724.00
1980	31340851	12824404	15091.00
1981	31246239	13209309	15542.00
1982	31459428	12716239	15982.00
1983	29692355	9366590.	16420.00
1984	29290353	7599711.	16862.00
1985	29346972	8097301.	17318.00
1986	31257561	8814210.	17790.00
1987	32377097	8860287.	18276.00
1988			

2.6. Guardar la vista como grupo

Para guardar la vista de las variables como grupo basta con dirigirse al menú de trabajo y hacer clic en “Name”, después de lo cual emergerá un cuadro de diálogo. En la sección “Name to identify object” se debe escribir el nombre con el cual se desea guardar la vista y finalmente se hace clic en “OK”.

2.7. Modificar datos y nombre de las series

Para editar los datos, ya sea de una serie individual o de una vista de series como grupo, se debe en el menú de trabajo, hacer clic en “Edit +/-”, después de lo cual aparecerá debajo del menú de trabajo una barra de modificación. Se debe hacer clic sobre el dato que se desea modificar, introducir la modificación y presionar ENTER.

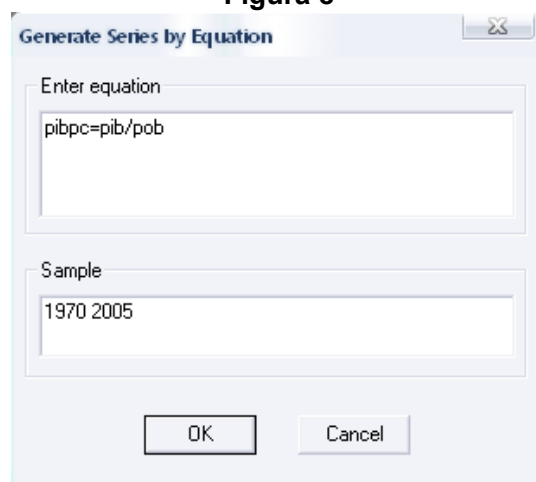
Para hacer cambios sobre el nombre de las variables, se debe llevar hasta arriba la barra de desplazamiento y, posicionándose sobre el nombre de la variable a editar, se introducen los cambios necesarios en el nombre de la variable y se presiona ENTER. Emergerá entonces un cuadro de diálogo, que pregunta: ¿desea renombrar la variable?, en donde se debe presionar “YES” si se está seguro de hacer los cambios y “NO” en caso contrario. Una vez realizados los cambios previstos, se debe dirigir de nuevo al menú de trabajo y hacer clic en “Edit +/-”.

2.8. Generar Series

Si entre las variables que se tienen están el producto interno bruto (PIB) y la población (POB), y se quiere por ejemplo transformar el PIB a PIB por habitante, el Eviews tiene una herramienta para hacerlo. La ruta desde el menú principal es: **Quick/Generate Series...**, después de lo cual emergerá un cuadro de diálogo como el que se muestra en la figura 8.

En la sección "Enter equation" se deberá introducir la ecuación correspondiente (primero el nombre de la variable a generar seguido del cálculo para obtener la serie solicitada) y, en "Sample" la muestra para la cual se desea realizar el cálculo. Es importante resaltar que, las variables involucradas en el cálculo se deben escribir como está su nombre en el archivo de trabajo.

Figura 8



Algunas de las fórmulas que se pueden utilizar en el Eviews son:

Operación	Forma de escribirla	Ejemplo
Logaritmo natural	LOG(X)	LCONS=LOG(CONS)
Función exponencial	EXP(X)	ECONS=EXP(CONS)
Valor absoluto	ABS(X)	VACONS=ABS(CONS)
Raíz cuadrada	SQR(X)	RCCONS=SQR(CONS)
Diferencia	D(X)	DIFERCONS=D(CONS)

2.9. Graficar Serie

Para graficar series, se debe abrir la variable contentiva de los datos que se desean graficar, luego, la ruta desde el menú de trabajo es: **View/Graph/Line** en caso de

querer un gráfico de línea y **View/Grpah/Bar** para gráficos de barras. Para volver a la vista de la hoja que contiene los datos, se debe en el menú de trabajo hacer clic el "Sheet".

2.10. Graficar Varias Series

Si se quieren graficar varias series simultáneamente, se deben abrir las series a graficar como grupo en el orden en el cual se desea aparezcan, luego se tienen dos opciones:

1ro. Graficar todas las series en un mismo grafico, para lo cual en el menú de trabajo se sigue la ruta: **View/Graph/Line** o **View/Graph/Bar**.

2do. Para graficar todas las series en gráficos separados pero simultáneos, la ruta desde el menú de trabajo es: **View/Multiple Graphs/Line** o **View/Multiple Graphs/Bar**.

Ahora bien, si se desea hacer un grafico de dispersión, basta con abrir como grupo las dos variables a graficar, presionando de primero la variable se desea que aparezca en el eje "X" y de segundo la variable se desea salga en el eje "Y". Luego, en el menú de trabajo se debe seguir la ruta: **View/Graph/Scatter/Simple Scatter** para un gráfico de dispersión simple o **View/Grpah/Scatter/Scatter with Regresión** para un grafico de dispersión con regresión (en esta segunda opción se pueden realizar algunas transformaciones a las variables según las necesidades del investigador). En el cuadro de diálogo emergente presionar "OK".

Otra opción surge desde el menú principal siguiendo la ruta: **Quick/Graph/...** para luego seleccionar el tipo de grafico que se desee y entonces, emergerá un cuadro de diálogo en donde se debe escribir la variable o grupo de variables a graficar, según sea el caso.

2.11. Guardar el Archivo de Trabajo

Para guardar el archivo de trabajo, se debe seguir la ruta desde el menú principal: **File/Save as**, después de lo cual emergerá un cuadro de diálogo, en donde se seleccionará la ubicación deseada para guardar el archivo de trabajo y luego de introducir el nombre hacer clic en "Guardar".

3. Primera Etapa del Proceso Econométrico:

Especificación del Modelo: Función de Consumo

Más allá de cualquier estudio anterior que haya sugerido la idea de alguna relación entre la renta y el ahorro agregado; la función de consumo es atribuida al economista británico J. M. Keynes, desarrollada en su obra *General Theory of Employment, Interest and Money hacia 1936*. En ella se expone que el consumo viene determinado por el ingreso en una relación lineal que puede ser descrita por dos propiedades, a saber, su inclinación como proporción del ingreso destinado al consumo, lo que Keynes denominó “propensión marginal a consumir” y, su ordenada de origen que representa aquel consumo que no depende del nivel de ingreso (consumo autónomo).

Otras teorías plantean la idea de que el consumo depende de la renta de largo plazo o permanente. Por un lado la hipótesis del ciclo vital considera que los individuos planifican el consumo y el ahorro con el fin de mantener un nivel de vida estable a lo largo del tiempo y, por el otro, la teoría de la renta permanente expone que se consumirá de acuerdo con el nivel actual de riqueza, el nivel de ingreso que se está percibiendo y una predicción del mismo a lo largo de toda la vida.

Se considere una u otra teoría, ciertamente, muchos coinciden en que la variable principal para explicar el consumo es el ingreso, ahora bien, se pueden presentar algunos problemas estadísticos desarrollados por Ackley (1965) y, que el investigador debe tomar en cuenta:

¿Ingreso real o ingreso monetario?

La proposición hecha por Keynes fue en términos de ingreso real, ya que, el efecto precios (inflación) impacta de manera significativa en el nivel de ingreso y específicamente en el poder adquisitivo del mismo, el cual se destinará finalmente al consumo. Además, desde el punto de vista econométrico, en busca de la parsimonia del modelo a estimar, ¿para qué incluir explícitamente la variable precios, es decir, como variable exógena, si puede estar inmersa de forma implícita proporcionando sencillez al modelo?.

¿Ingreso total o ingreso disponible?

Pensar que lo que cuenta es el ingreso disponible es suponer, como se hace usualmente, que una reducción en los impuestos de un hombre afectará su consumo de la misma manera que un aumento equivalente en su ingreso posterior al pago de impuesto surgido de ingresos más altos antes del pago de impuesto.

¿Ingreso agregado o ingreso *per cápita* (pc)?

Keynes aseguraba que los resultados agregados, eran consecuencia de la suma de las acciones individuales. Sin embargo, una serie corregida por el crecimiento poblacional, proporcionará mejores resultados ya que, un aumento del ingreso como consecuencia del crecimiento poblacional con el mismo ingreso promedio *per cápita*, tendrá resultados diferentes si se compara con un crecimiento del ingreso agregado con una población estable y un ingreso *per cápita* en aumento.

Por otro lado, existen algunas evidencias empíricas acerca de la función de consumo de largo plazo donde si existe una tendencia en la misma, basta incluir como lo hizo Smithies la variable tiempo, que recogerá el comportamiento ascendente o descendente (los datos estaban corregidos por los precios y la población). La posibilidad de que haya un retardo en el consumo, el tipo de interés, el volumen de la riqueza, y la existencia de dinero, son algunas otras influencias sugeridas por diversos autores sobre el gasto en consumo.

En resumen se puede especificar un modelo para el consumo de la siguiente manera:

$$C = f(Yd, Pob, IPC, T, C_{t-1}, i, w, L)$$

donde, C es el gasto en consumo, Yd es el ingreso disponible, Pob es la población, IPC es el índice de precios al consumidor, T es el tiempo, i es la tasa de interés, w es el volumen de riqueza y L es la existencia de dinero. No obstante, partiendo de los términos más básicos, se partirá de estimar un modelo simple de regresión lineal para el caso venezolano del consumo como se expone a continuación:

$$C_{pc} = \beta_0 + \beta_1 Yd_{pc} + \mu \quad (1)$$

donde, β_1 es la propensión marginal a consumir (PMC), de la cual se espera un signo positivo entre cero y uno. β_0 , es el consumo autónomo, pero como el análisis que se pretende en este documento es sobre la PMC no es relevante el signo que arroje la estimación (tradicionalmente es positivo). Los datos deberán estar corregidos por el crecimiento de los precios y la población.

Ahora bien, sería poco lógico afirmar que todos los individuos tienen la misma PMC, sin embargo, estimar dicha propensión para cada individuo resulta utópico; es por ello que, se pueden definir en general dos tipos de personas, los dueños del factor capital y los dueños del factor trabajo, de lo cual surge un segundo modelo:

$$C = \beta_0 + \beta_1 YKd + \beta_2 YLd + \beta_3 Pob + \mu \quad (2)$$

donde, YKd es la remuneración a los dueños del factor capital y YLd es la remuneración a los dueños del factor trabajo y, aunque los datos deben estar corregidos por los precios, no se pueden trabajar en términos *per cápita* ya que, dividir estas remuneraciones por la población no resulta lógico, de allí que la población (Pob) sea una variable exógena adicional en el modelo. Los resultados esperados, es que la PMC_K sea menor que la PMC_L , ya que éstos tienen menor ingreso que aquellos y, a menor ingreso, mayor será la proporción destinada al consumo, por su lado, a mayor ingreso, mayor será la propensión marginal a ahorrar. Ambos valores deben ser positivos y estar entre cero y uno. Por su lado, β_3 debe ser positivo indicando que a mayor crecimiento poblacional mayor será el consumo. Los datos deberán estar corregidos por el crecimiento de los precios.

Se pretende además en este documento, analizar el impacto de algunas otras variables que puedan influir sobre el gasto en consumo en Venezuela como se mencionó antes.

4. Segunda Etapa del Proceso Econométrico:

Recolección de Datos

Los datos recolectados ya sea de fuentes primarias (obtención propia, por ejemplo las encuestas) o secundarias (datos reelaborados, analizados o sintetizados por otros, por ejemplo estadísticas del Banco Central de Venezuela), deben ser, precisos (recolectados en forma confiable), pertinentes (representativos del objeto de estudio) y consistentes (sujetos a una misma pauta de recolección). En general, los economistas recurren a fuentes secundarias para obtener datos; a continuación se presentan algunas instituciones de reconocido prestigio con sus respectivas páginas web donde se pueden encontrar datos que cumplan con las características ya expuestas:

Banco Central de Venezuela	www.bcv.org.ve
Instituto Nacional de Estadística	www.ine.gob.ve
Comisión Económica para América Latina y el Caribe	www.eclac.org
Fondo Monetario Internacional	www.imf.org
Banco Mundial	www.bancomundial.org

Antes de pasar a la siguiente etapa del proceso econométrico es importante hacer un análisis descriptivo de las series a utilizar, a saber, estadística descriptiva, tasas de crecimiento (en caso de series de tiempo), participaciones del total (para datos de corte transversal), interpretación de gráficos y de ser posible análisis comparativo que evidencien hallazgos de relevancia para la investigación.

En el caso que nos ocupa, hacia el desarrollo de este documento, los datos fueron extraídos de las Bases Cuantitativas de la Economía Venezolana 1830-2002, publicado por el profesor Asdrúbal Baptista en 2006; y la muestra establecida fue desde 1950 hasta 2002.

4.1. Estadística Descriptiva

Para calcular la estadística descriptiva de los datos, se deben abrir las series objeto del cálculo como grupo, luego, desde el menú de trabajo seguir la ruta: **Descriptive Stats/Common Sample** (para datos comunes en las series seleccionadas), después de lo cual el Eviews arrojará la estadística descriptiva con los siguientes elementos:

	CONSUMOPC	INGRESODPC
Mean	13721.43	16345.54
Median	13589.09	15657.45
Maximum	23173.04	28302.97
Minimum	5703.665	7298.405
Std. Dev.	4788.654	5694.464
Skewness	0.168855	0.612994
Kurtosis	2.153516	2.716013
Jarque-Bera Probability	1.834204 0.399676	3.497322 0.174007
Sum	727235.7	866313.7
Sum Sq. Dev.	1.19E+09	1.69E+09
Observations	53	53

Mean (Media): que desde el punto de vista de la Física, la media de una serie de datos representa el centro de gravedad o punto de equilibrio de esos datos.

Median (Mediana): es aquel valor que ocupa la posición central de los datos, una vez que estos han sido ordenados.

Maximum (Valor Máximo): es el valor máximo de la serie bajo estudio.

Minimum (Valor Mínimo): es el valor mínimo de la serie bajo estudio.

Std. Dev. (Desviación Estándar): mide la dispersión de los datos con respecto a su media; de lo cual se deduce que mientras más alejados estén los valores de su media, entonces mayor será el valor de la desviación y mientras más concentrados se encuentren alrededor de su media, menor será el valor de la desviación estándar.

Skewness (Asimetría): una distribución es simétrica respecto a un punto, cuando las frecuencias se parten siguiendo el mismo patrón a ambos lados de ese punto. Cuando una distribución no es simétrica, se dice que es asimétrica u oblícua. La asimetría puede ser positiva o por la derecha, cuando la mayoría de los datos se concentran hacia la izquierda; negativa o por la izquierda, si los datos se ubican mayormente hacia la derecha. Asimetría de una distribución unimodal:

- =0, si es simétrica
- >0, si es asimétrica por la derecha
- <0, si es asimétrica por la izquierda

Kurtosis (Curtosis): es una medida de apuntamiento que proporciona una idea del grado en que la curva correspondiente a una distribución de frecuencia es apuntada o achatada en su parte central, es decir, mide el nivel de concentración de datos en la región central. Aunque no reflejan lo mismo, la curtosis también puede describirse como un grado de dispersión. La medición del achatamiento de una curva se hace mediante un patrón de referencia que es la Curva Normal o Curva de Gauss, es decir, una curva unimodal simétrica con forma de campana. A esta curva se le denomina mezocurtica. Cuando la curva es más puntiaguda que la normal se le llama leptocurtica y cuando es más achatada se le denomina platicurtica.

Jaque-Bera: es un estadístico distribuido con un χ^2 con dos grados de libertad, que permite verificar si la serie está o no normalmente distribuida.

Probability: probabilidad asociada al estadístico Jaque-Bera.

Observations (Observaciones): es el número de observaciones incluidas en la serie.

4.2. Tasa de Crecimiento

Para conocer la tasa de crecimiento promedio de una serie de tiempo, basta con calcular el crecimiento año tras año, a través de la siguiente fórmula:

$$r = \left(\frac{X_n}{X_1} \right)^{\frac{1}{n}} - 1$$

donde, X_n es el valor final y X_1 el valor inicial, ambos del período sujeto al cálculo, y n , es el número de períodos que para el caso de años consecutivos será uno. Luego, se debe obtener un promedio simple éstos datos. No obstante, una vez estudiado el análisis de regresión, podremos obtener la tasa de crecimiento a partir del mismo.

4.3. Participación del Total

En caso de que se estén utilizando datos de corte transversal en muestras pequeñas, para enriquecer la descripción de los datos, es recomendable calcular la participación de cada observación dentro del total, con el fin de comparar quiénes tienen mayor o menor peso sobre la variable objeto de estudio. Por ejemplo y utilizando valores hipotéticos, si tenemos el valor del consumo para 2007 de un conjunto de países, la forma de calcular la participación del total sería como se muestra a continuación:

País	Consumo	Fórmula de cálculo	Participación del Total
Venezuela	80	80/280	28,6%
Colombia	30	30/280	10,7%
Argentina	50	50/280	17,9%
Brasil	120	120/280	42,8%
Total	280	280/280	100%

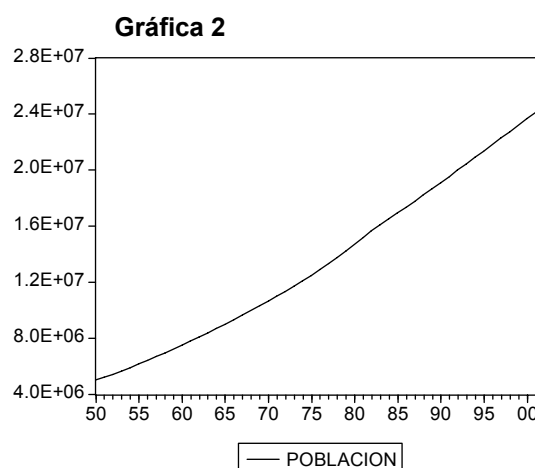
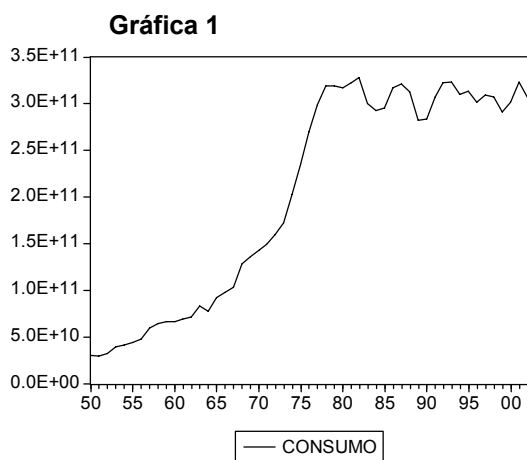
La fórmula general para calcular la participación de una observación sobre el total es la siguiente:

$$\% \text{ de participación} = \frac{\text{Valor de la observación}}{\text{Valor total}} * 100$$

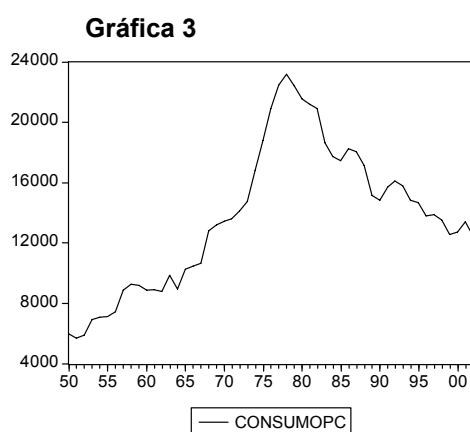
4.4. Análisis Descriptivo de las Variables: CONSUMO e INGRESOd

Para entender la corrección de las variables por la población, se debe comenzar por describir las mismas en su valor agregado. El valor promedio del consumo total para Venezuela durante el período 1950-2002, fue de 202.627.104.175 bs de 1984, tal que; el valor mínimo fue de 28.841.406.238 bs de 1984 en 1951 y el valor máximo ocurrió en 1982 tomando un valor de 328.105.654.988 bs de 1984. Al mismo tiempo, la tasa de crecimiento promedio interanual fue de 4,83%, sin embargo; al observar la gráfica 1, se puede en efecto detallar un crecimiento sostenido del consumo total desde 1950 hasta 1979, en contraste con el resto del período de estudio, en donde la variable fluctúa sin una tendencia clara y significativa. Ésta diferencia en cuánto al comportamiento del consumo total, antes y después de 1979, se puede recoger estimando un modelo de tasa de crecimiento con una variable dicotómica que tome en cuenta el efecto intercepto y el efecto pendiente, no obstante, esto se estudiará más adelante.

Por su parte, la población venezolana ha tenido un comportamiento creciente y sostenido durante el período de estudio, como se pone de manifiesto en la gráfica 2; con una tasa de crecimiento promedio interanual de 3,06%. Ahora bien, si se corrige el consumo total por el crecimiento poblacional; esto impactará de manera significativa en el comportamiento del consumo; sobre todo en el período comprendido entre 1980 y 2002, hecho que se evidencia al observar la gráfica 3. Ciertamente, con un consumo total que fluctúa en torno a una pendiente que se aproxima a cero, dividido por el crecimiento poblacional, marca una clara tendencia decreciente durante el período 1980-2002 para el consumo *per cápita* de Venezuela. De allí la importancia de transformar las variables en términos *per cápita*.



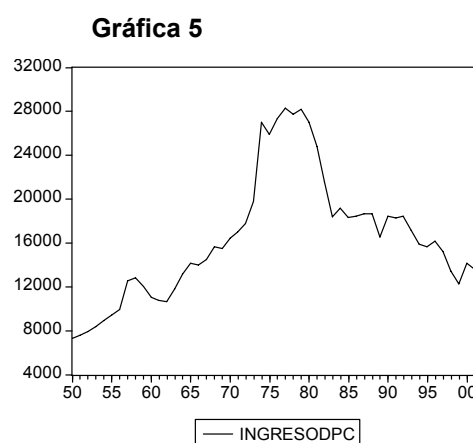
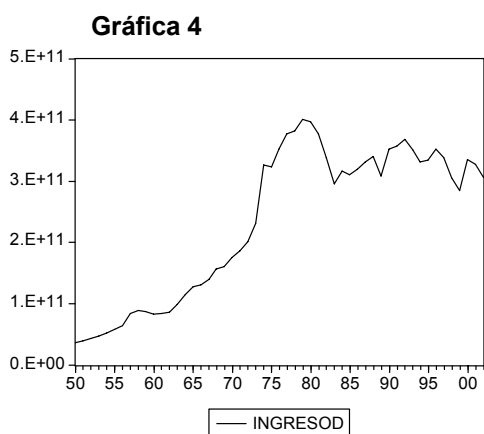
El valor promedio del consumo *per cápita* de Venezuela durante el período 1950-2002 fue de 13.721,43 bs de 1984; alcanzando un valor máximo de 23.173,04 bs de 1984 en 1978 y un valor mínimo de 5.703,67 bs de 1984 en 1951, lo que se tradujo en una desviación estándar de 4.788,65 bs de 1984. La tasa de crecimiento promedio interanual del consumo *per cápita* fue de 1,68% para el mismo período de estudio, sin embargo, con la utilización de variables dicotómicas, más adelante, se podrá hacer la distinción entre la tasa de crecimiento del período 1950-1979 donde el consumo *per cápita* tiene una tendencia creciente y 1980-2002 donde la variable decrece.



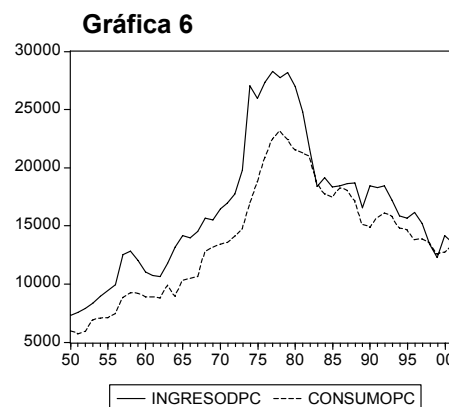
El ingreso disponible de Venezuela entre 1950-2002, por su parte, alcanzó un valor promedio de 234.398.328.147 bs de 1984 y; su valor máximo fue de 401.286.646.018 bs de 1984 en 1979, mientras que el valor mínimo ocurrió en 1950 llegando a los 36.746.284.374 bs de 1984. La tasa de crecimiento promedio interanual fue

4,23%, como se puede observar en la gráfica 4, aunque, es de hacer notar, al igual que ocurre con el consumo total que a partir de 1979 cambia de manera significativa su comportamiento.

Al corregir los valores del ingreso disponible por el crecimiento poblacional se obtiene que, el valor promedio del ingreso disponible *per cápita* de Venezuela durante el período de estudio fue de 16.345,54 bs de 1984, tal que, el valor mínimo fue de 7.298,4 bs de 1984 en 1950 y el valor máximo fue de 28.302,97 bs de 1984 en 1977, lo que significó una desviación estándar de 5.694,46 bs de 1984. Al mismo tiempo, la tasa de crecimiento promedio interanual fue de 1,16% (véase gráfica 5).



Finalmente, haciendo un análisis comparativo entre el consumo y el ingreso disponible, ambos *per cápita*, se puede percibir una estrecha correlación entre ambos como se pone de manifiesto en la gráfica 6. Ambas con una tasa promedio creciente durante el período de estudio, sin embargo, con un punto de inflexión notorio hacia 1979.



Nota: para hacer un análisis más completo y preciso de las variables objeto de estudio, hace falta conocer algunas herramientas que el análisis de regresión proporcionará.

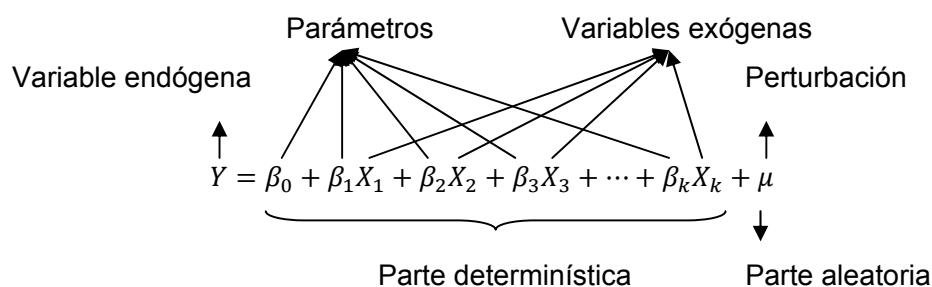
5. Tercera Etapa del Proceso Econométrico:

Estimación del Modelo

5.1. Análisis de Regresión

Estudia el grado de dependencia de una variable llamada endógena, con respecto a una o más variables, conocidas como exógenas; con el fin de estimar y/o predecir el valor medio de la primera en términos de los valores conocidos o fijos de las segundas. En otras palabras, el análisis de regresión, busca describir y evaluar, a través de la estimación de parámetros, la relación que existe entre una variable endógena (Y) y un conjunto de variables independientes ($X_1, X_2, X_3, \dots, X_n$).

Dichos parámetros pueden ser estimados a través de varios métodos, tomando en cuenta si el modelo es uniecuacional o multiecuacional, lineal o no lineal. En este caso, se está hablando de un modelo lineal, uniecuacional, multiple; es decir, un Modelo Clásico de Regresión Lineal (MCLR) representado de la siguiente manera:



Para estimar los valores de los parámetros de este tipo de modelos, se pueden utilizar los siguientes métodos: Mínimos Cuadrados Ordinarios (MCO), Mínimos Cuadrados Indirectos, Mínimos Cuadrados en dos etapas, Máxima Verosimilitud, entre otros. En el caso que nos ocupa, se utilizará el método de los Mínimos Cuadrados Ordinarios, atribuido a Carl Friedrich Gauss; método que obtiene unos parámetros con ciertas propiedades que los hacen los Mejores Estimadores Lineales Insesgados (MELI), además, es sencillo y no requiere tamaños de muestras grandes.

Es importante resaltar que detrás del MCRL y el método de los MCO, existen una serie de supuestos que deben ser cumplidos al momento de realizar la estimación, para asegurar que los estimadores sean MELI; a saber:

- 0ro. El modelo es verdadero y existe.
- 1ro. Es un modelo de regresión lineal (lineal en los parámetros)
- 2do. Los valores de las variables exógenas son fijos en muestreos repetidos.
- 3ro. El valor medio de los residuos es igual a cero.
- 4ro. La varianza de los residuos es homocedástica.
- 5to. Los residuos no están autocorrelacionados.
- 6to. La covarianza entre los residuos y las variables exógenas es igual a cero.
- 7mo. El número de observaciones debe ser mayor que el número de parámetros a estimar.
- 8vo. Los valores de las variables exógenas deben variar.
- 9no. El modelo está correctamente especificado.
- 10mo. No existe multicolinealidad perfecta entre las variables exógenas.
- 11vo. Los residuos están distribuidos normalmente.

5.2. Estimación

Se comenzará por estimar un modelo simple de regresión lineal como se expuso en su momento de la siguiente manera:

$$Cpc = \beta_0 + \beta_1 Ydpc + \mu \quad (1)$$

Variable endógena: consumo *per cápita* (Cpc), medido en bolívares de 1984 y definido como el gasto destinado a la compra de bienes y servicios.

Variable exógena: ingreso disponible *per cápita* (Ydpc), medido en bolívares de 1984 y definido como la remuneración en unidades monetarias por actividades o no productivas.

Parámetros a estimar: β_0 interpretado como el consumo que no depende del nivel de ingreso y β_1 es la propensión marginal a consumir (PMC), la cual debe ser positiva y estar entre cero y uno.

Para realizar un análisis de regresión, desde menú principal la ruta a seguir es: **Quick/Estimate Equation...**, luego emergerá un cuadro de diálogo como el que muestra la figura 9. En la sección de **“Equation specification”** se debe introducir el nombre de la variable dependiente o endógena, seguido de una “C” (letra que reserva el Eviews para el intercepto) y a continuación las variables independientes o exógenas; todas separadas por un espacio. En el apartado **“Method”** se debe seleccionar el método de estimación que por defecto es Mínimos Cuadrados Ordinarios (**Least Squares**) y en la sección **“Sample”** se debe introducir el tamaño de la muestra. Finalmente se presiona **“OK”** y se tendrá la salida de la estimación que se pueden detallar en la figura 10.

Figura 9

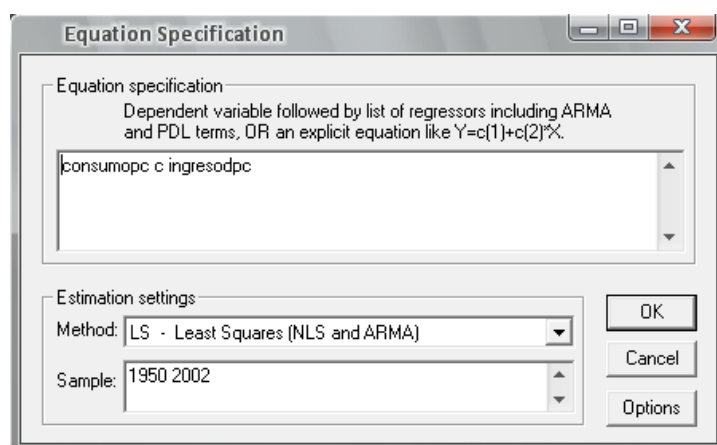


Figura 10

Dependent Variable: CONSUMOPC				
Method: Least Squares				
Date: Time:				
Sample: 1950 2002				
Included observations: 53				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	820.2701	702.6204	1.167444	0.2485
INGRESODPC	0.789277	0.040634	19.42398	0.0000
R-squared	0.880922	Mean dependent var		13721.43
Adjusted R-squared	0.878587	S.D. dependent var		4788.654
S.E. of regression	1668.575	Akaike info criterion		17.71433
Sum squared resid	1.42E+08	Schwarz criterion		17.78868
Log likelihood	-467.4298	F-statistic		377.2912
Durbin-Watson stat	0.401996	Prob(F-statistic)		0.000000

Los elementos que se distinguen en la salida de la estimación son:

Dependent Variable: variable dependiente.

Method: Least Squares (LS): método de estimación: mínimos cuadrados ordinarios.

Date – Time: fecha y hora de la estimación.

Sample: muestra tomada en cuenta para la estimación.

Observations: número de observaciones.

Variable: nombre de las variables independientes.

Coefficient (coeficientes): recoge el valor de los parámetros asociados a cada una de las variables explicativas.

Std. Error (Error Estándar): recoge la desviación típica estimada de los estimadores y nos indica el grado de confianza que podemos depositar en nuestra estimación.

t-statistic (t-estadístico): se calcula como el cociente entre el estimador y su error estándar y nos permite contrastar hipótesis.

Prob. (p-valué): indica la probabilidad de cometer el error de tipo I (rechazar la hipótesis nula siendo cierta).

R-squared (R^2): es el coeficiente de determinación y valora el éxito de la regresión.

Adjusted R-squared (R^2 ajustado): se obtiene a partir del coeficiente de determinación ponderándolo por los grados de libertad.

S.E. of regresión (Suma de los Errores al Cuadrado): es el valor de la función objetivo en el mínimo, cuando estimamos por mínimos cuadrados ordinarios.

Log likelihood (logaritmo de la función de verosimilitud): es el valor de la función objetivo en el máximo, cuando estimamos por máxima verosimilitud.

Durbin-Watson stat (Estadístico Durbin-Watson): sirve para contrastar la autocorrelación.

Mean dependent var: media de la variable dependiente.

S.D. dependent var: cuasi desviación típica muestral de la variable dependiente.

Akaike info criterion: criterio de información Akaike.

Schwarz criterion: criterio Schwarz.

Éstos dos últimos criterios miden la capacidad explicativa de un modelo y permiten realizar comparaciones a este respecto entre modelos anidados.

F-statistic (estadístico F): sirve para contrastar si los parámetros asociados a las variables explicativas del modelo son conjuntamente igual a cero.

Prob(F-statistic) (p-value): mide la probabilidad de cometer el error de tipo I.

5.3. Representación del Modelo

Para saber como se representa el modelo estimado, la ruta a seguir desde el menú de trabajo es: **View/Representations** y arrojará una salida como la que se muestra en la figura 11; en donde se puede observar el método de estimación y las variables involucradas en el modelo (*Estimation Command*), el modelo a estimar (*Estimation Equation*) y finalmente el modelo estimado (*Substituted Coefficients*). Para volver a la salida del modelo, basta con hacer clic sobre **Stats** en el menú de trabajo.

Figura 11

```

Estimation Command:
=====
LS CONSUMOPC C INGRESODPC

Estimation Equation:
=====
CONSUMOPC = C(1) + C(2)*INGRESODPC

Substituted Coefficients:
=====
CONSUMOPC = 820.2700605 + 0.7892769433*INGRESODPC
    
```

La manera correcta de escribir el modelo estimado es:

$$\hat{C}_{pc} = 820,27 + 0,79Ydpc \quad (1.1)$$

$$t \quad (1,1674) \quad (19,4239)$$

$$R^2 = 0,8809 \quad F = 377,29$$

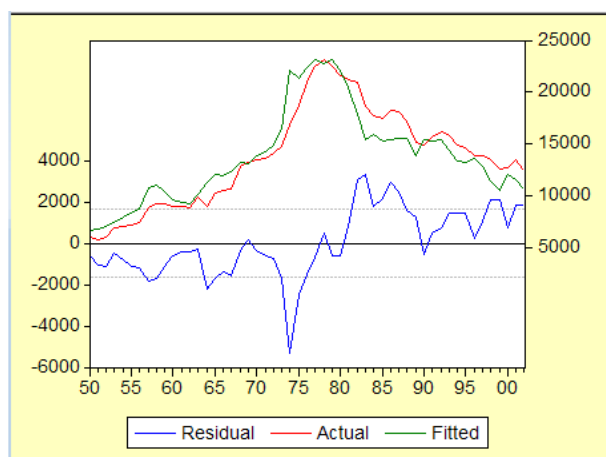
Nótese que la diferencia entre el modelo especificado y el estimado, es que éste último ya no tiene la parte aleatoria y sobre la variable endógena se escribe un “gorro” que significa estimado. Por su lado, los parámetros estimados satisfacen lo planteado en la primera etapa del proceso econométrico; la PMC está entre cero y uno, al mismo tiempo que el intercepto es positivo.

5.4. Grafica del Modelo

Una vez estimado el modelo, para saber como ha sido el ajuste de la línea de regresión a los valores verdaderos del consumo *per cápita*, desde el menú de trabajo la ruta a seguir es: **View/Actual,Fitted,Residual/Actual,Fitted,Residual Graph** o haciendo clic en **Resids** en el menú de trabajo. El Eview mostrará una gráfica de los verdaderos valores del consumo *per cápita* (*Actual*), el valor estimado del consumo *per cápita* (*Fitted*) y la diferencia entre ambos, es decir, los residuos o errores (*Residual*).

En ella se deben observar dos cosas, la primera que la línea estimada (de color verde) se aproxime lo más como sea posible a la línea observada (de color rojo) y segundo, que los residuos se comporten de manera estocástica o aleatoria (véase figura 12).

Figura 12



5.5. Valores observados, estimados y residuos

Para conocer los valores observados, estimados y de los residuos, la ruta desde el menú de trabajo es: **View/Actual,Fitted,Residual/Actual,Fitted,Residual Table**, y entonces se obtendrán los valores que se graficaron en la figura 12, además se tendrá también la grafica de los residuos (*Residual Plot*).

5.6. Modelo Múltiple de Regresión Lineal

El paso de un modelo simple de regresión lineal a uno múltiple es muy sencillo, basta sólo con incluir el resto de las variables exógenas para que el Eviews realice la estimación. Para ilustrarlo se debe recordar el segundo modelo especificado en la etapa uno del proceso econométrico, a saber:

$$C = \beta_0 + \beta_1 YKd + \beta_2 YLd + \beta_3 Pob + \mu \quad (2)$$

Variable dependiente: consumo total, medido en bolívares de 1984.

Variables independientes: ingreso disponible de los dueños del factor capital (YKd), ingreso disponible de los dueños del factor trabajo (YLd), ambos medidos en bolívares de 1984 y la población medida en personas.

Parámetros a estimar: β_0 que es el consumo que no depende del nivel de ingreso, β_1 PMCK, β_2 PMCI y β_3 que recogerá el impacto que sobre el consumo tiene el crecimiento de la población.

Nuevamente la ruta a seguir desde el menú principal es: **Quick/Estimate Equation...**, luego, en el cuadro de diálogo emergente se debe repetir el procedimiento explicado en la sección estimación de este documento sólo que se deben escribir tantas variables exógenas como lo indique el modelo especificado.

5.7. Guardar la Estimación

Para guardar la estimación realizada, se debe en el menú de trabajo hacer clic sobre **Name**, después de lo cual aparecerá un cuadro de diálogo y en la sección "Name to indentify object" escribir el nombre deseado para luego hacer clic en "OK".

5.8. Modificar la Estimación

Una vez abierta la estimación, si se desea realizar algún cambio, ya sea en las variables, el método o la muestra de la misma; desde el menú de trabajo se debe hacer clic en "Estimate" y emergerá un cuadro de diálogo como el mostrado en la figura 9 correspondiente a la estimación abierta. Una vez hechos los cambios, se debe hacer clic en "OK" para aceptar los cambios introducidos.

5.9. Relaciones Funcionales

Algunas de las razones para utilizar algunas relaciones funcionales, primero para suavizar el comportamiento de alguna variable en especial y para ello por ejemplo se aplica logaritmo, segundo para estimar una función en específico por ejemplo una Coob-Douglas (Log-Log), tercero para recoger algún impacto en específico como la tasa de crecimiento, entre otras. Las formas funcionales más comunes que se pueden estimar son las siguientes:

Relación Funcional		Línea a Modelar	Interpretación de β_i
Lin-Lin	Relación Lineal	$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$	Por cada unidad que aumenta X_1 , Y aumenta en β_i unidades, durante el período "XX", manteniendo las demás variables constantes
Log-Log	Elasticidad	$\text{Log}(Y) = \beta_0 + \beta_1 \text{Log}(X_1) + \beta_2 \text{Log}(X_2) + \mu$	Por cada 1% que aumenta X_1 , Y aumenta en $\beta_i\%$, durante el período "XX", manteniendo las demás variables constantes
Log-Lin	Cuando la variable exógena es el tiempo* (Tasa de crecimiento)	$\text{Log}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \mu$	Si X_1 =Tiempo: La tasa de crecimiento promedio interanual de Y fue de $(\beta_i * 100)\%$, durante el período "XX"
Lin-Log	Semi-elasticidad	$Y = \beta_0 + \beta_1 \text{Log}(X_1) + \beta_2 \text{Log}(X_2) + \mu$	Por cada 1% que aumenta X_1 , Y aumenta en $(\beta_i/100)$ unidades, durante el período "XX", manteniendo las demás variables constantes

* Para crear una serie llamada tiempo o tendencia, desde la ruta *Quick/Empty Group (Edit Series)* se debe escribir en la ventana emergente los números del 1 hasta "n", donde "n" representa el número de observaciones incluidas en el modelo. En caso de que los datos sean anuales, utilizar el valor del año en cuestión también sirve, ahora bien, de utilizar una u otra forma implican cambios sólo en el intercepto y no en el estimador que recoge la tasa de crecimiento.

Así por ejemplo, si se desea estimar una **relación lineal** (Lin-Lin), es decir, con las variables en su estado natural (serán validas las transformaciones económicas más no logarítmicas), basta con introducir las variables como se hizo en el apartado 5.2. de este documento y de ser un modelo múltiple adicionar las variables restantes, para obtener:

$$\hat{C}_{pc} = 820,27 + 0,79Y_{dpc} \quad (1.1)$$

Ahora bien, si se quiere estimar una función de producción de tipo Coob-Douglas, entiéndase que es una relación funcional no lineal de la siguiente manera:

$$Q = AK^{\beta_1}L^{\beta_2}\varepsilon \quad (3)$$

donde Q son las cantidades producidas, K las unidades de factor capital, L las unidades del factor trabajo, A un término constante, ε el término de perturbación aleatoria y β_i los parámetros a estimar. Para poder obtener unos estimadores mínimos cuadrados es necesario linealizar esa función de producción a través de la transformación logarítmica como se hace a continuación:

$$\text{Log}(Q) = \text{Log}(AK^{\beta_1}L^{\beta_2}\mu) \quad (3.1)$$

$$\text{Log}(Q) = \text{Log}(A) + \text{Log}(K)^{\beta_1} + \text{Log}(L)^{\beta_2} + \text{Log}(\mu) \quad (3.2)$$

$$\text{Log}(Q) = \text{Log}(A) + \beta_1\text{Log}(K) + \beta_2\text{Log}(L) + \text{Log}(\mu) \quad (3.3)$$

$$\text{Log}(Q) = \alpha_0 + \beta_1\text{Log}(K) + \beta_2\text{Log}(L) + \varepsilon \quad (3.4)$$

de esta forma se obtiene una **relación funcional con coeficientes de elasticidad** (Log-Log) y, los parámetros estimados recogerán el impacto que un aumento del uno por ciento (1%) en la variable exógena tiene en términos también porcentuales sobre la variable endógena.

En caso de que se desee estimar la **tasa de crecimiento** (Log-Lin) de una variable que necesariamente debe ser una serie de tiempo, basta con colocar la variable objeto de estudio en función del tiempo de la siguiente manera:

$$\text{Log}(C_{pc}) = \beta_0 + \beta_1\text{Tiempo} + \mu \quad (4)$$

obteniéndose la siguiente recta de regresión:

$$\text{Log}(\hat{C}_{pc}) = -23,85147 + 0,016858\text{Tiempo} \quad (4.1)$$

El parámetro ($\hat{\beta}_1 * 100$) recoge la tasa de crecimiento promedio interanual del consumo *per cápita*; éste valor fue utilizado en el análisis descriptivo de las variables en la segunda etapa del proceso econométrico.

Otro caso se puede presentar, si por ejemplo, con datos de corte transversal se quiere recoger el efecto que un aumento del uno por ciento (1%) del ingreso tiene sobre el nivel de consumo, caso en que es necesaria una estimación de **semi-elasticidad** (Lin-Log), como a continuación se aplica:

$$C_{pc} = \beta_0 + \beta_1 \text{Log}(Yd_{pc}) + \mu \quad (5)$$

Sin embargo, esto no quiere decir que éstas relaciones funcionales se den de manera aislada, por el contrario, son posibles también las combinaciones entre ellas. Un ejemplo de ello se explica a continuación:

Considérese el primer modelo especificado en la ecuación (1) acerca del consumo ($C_{pc} = \beta_0 + \beta_1 Yd_{pc} + \mu$), no obstante en el largo plazo como se expuso en el análisis descriptivo del comportamiento del consumo *per cápita*, éste ha crecido a una tasa promedio interanual de 1,68%. Por lo tanto, siguiendo a Smithies, incluir la variable tiempo en el modelo recogerá la tendencia creciente o decreciente del consumo, con los datos corregidos por la población y los precios como en efecto están. Vale resaltar que, para el tiempo capte el efecto planteado con anterioridad, habrá que aplicar logaritmo al consumo *per cápita*, dando lugar a un nuevo modelo:

$$\text{Log}(C_{pc}) = \beta_0 + \beta_1 Yd_{pc} + \beta_2 \text{Tiempo} + \mu \quad (6)$$

Arrojando los siguientes resultados:

$$\text{Log}(\hat{C}_{pc}) = -9,91 + 5E^{-5}Yd_{pc} + 0,009\text{Tiempo} \quad (6.1)$$

Nótese que la propensión marginal a consumir es tan pequeña que resulta no significativa en términos económicos, es por ello que replanteando el modelo se debe estimar un coeficiente de elasticidad en la relación ingreso-consumo y un coeficiente de tasa de crecimiento para la relación tiempo-consumo de la siguiente manera:

$$\text{Log}(C_{pc}) = \beta_0 + \beta_1 \text{Log}(Yd_{pc}) + \beta_2 \text{Tiempo} + \mu \quad (7)$$

de lo que se obtiene:

$$\text{Log}(\hat{C}_{pc}) = -12,27 + 0,87\text{Log}(Y_{dpc}) + 0,006\text{Tiempo} \quad (7.1)$$

Es de hacer notar como, éste ultimo modelo estimar que involucra una relación Log-Log y Log-Lin, es más coherente con la teoría económica, de allí que las relaciones funcionales sean tan útiles para recoger efectos específicos, además de ser un método de suavización como más adelante se mencionará.

6. Cuarta Etapa del Proceso Econométrico:

Evaluación del Modelo: Función de Consumo

6.1. Evaluación Económica: consiste en verificar si los signos y las magnitudes de los parámetros estimados satisfacen la teoría económica, para lo cual se tomará como referencia el primer modelo estimado:

$$\begin{aligned} \hat{C}_{pc} &= 820,27 + 0,79Y_{dpc} && (1.1) \\ t & \quad (1,1674) \quad (19,4239) \\ R^2 &= 0,8809 \quad F = 377,29 \end{aligned}$$

Queda a juicio del investigador justificar con fundamentos teóricos si los signos y las magnitudes satisfacen la teoría económica. En el caso de $\hat{\beta}_1$, el parámetro satisface los criterios establecidos por Keynes en cuanto a su planteamiento acerca de la función de consumo, el cual explicaba que ése valor es la proporción del ingreso destinada al consumo de bienes y servicios por lo que debe estar entre cero y uno. Por su lado, $\hat{\beta}_0$ interpretado como aquél consumo que no depende del nivel de ingreso (consumo autónomo) debe ser positivo, lo que en efecto ocurre.

De acuerdo a lo antes señalado, se concluye que el modelo es económicamente significativo, por lo que se procede a la evaluación estadística.

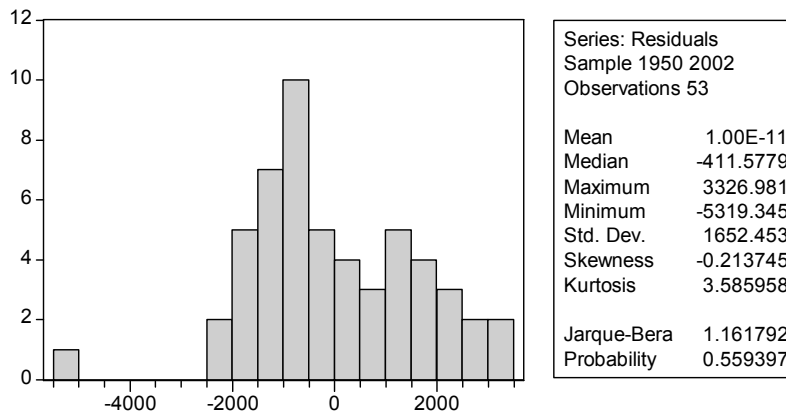
Nota: El supuesto de normalidad de los residuos: antes de entrar a la evaluación estadística, es importante hacer una parada en el supuesto de normalidad de los residuos. Aunque la aplicación del método del MCO para estimar un MCRL no requiere ningún supuesto acerca de la distribución de probabilidad de los residuos; el objetivo de objetivo del análisis de regresión es estimar una función de regresión muestral para hacer inferencia acerca de la función de regresión poblacional, lo que significa aplicar test de hipótesis sobre los parámetros, para lo que es necesario especificar la distribución de probabilidad de los residuos, ya que, los parámetros son una función de aquellos.

Asumir que $\mu \sim N(0, \sigma^2)$, trae ciertas propiedades sobre los estimadores MCO, a saber, son insesgados, varianza mínima (eficientes) y consistentes, es decir, $\hat{\beta}_i \rightarrow \beta_i$ a medida que el número de observaciones aumenta.

Para contrastar la normalidad de los residuos, se debe aplicar el **test de Jaque-Bera** siguiendo la ruta desde el menú de trabajo de la estimación: **View/Residual test/Histogram – Normality test**, para luego realizar el procedimiento que a continuación se expone:

Planteamiento de hipótesis	Hipótesis nula: $H_0: \mu \sim N(0, \sigma^2)$ Hipótesis alternativa: $H_1: \mu \text{ No } \sim N(0, \sigma^2)$
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	$JB = n \left[\frac{S^2}{6} + \frac{(K - 3)^2}{24} \right] \sim X^2$ donde, S es el coeficiente de asimetría y K es el coeficiente de cortosis.
Regla de decisión	Si $\alpha > p\text{-valué}$, rechazo H_0 , con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que los residuos (<u>se o no se</u>) distribuyen normal.

En el modelo estimado en la ecuación (1.1) los resultados que el test de Jarque-Bera arroja son:



Regla de decisión: $(\alpha = 0,05) < (p - \text{valué} = 0,56)$. Además nótese que el coeficiente de asimetría se aproxima a cero y el de curtosis a tres (véase su significado en la estadística descriptiva).

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5%, que los residuos se distribuyen normal.

Las consecuencias de la no normalidad de los residuos son dos, la primera, los estimadores MCO siguen siendo MELI bajo los supuestos del MCRL, la segunda, las conclusiones obtenidas a partir de las pruebas de hipótesis no son válidas ya que si los residuos no son normales, los estimadores MCO tampoco lo son.

La posible solución a la no normalidad es, acudir al teorema central del límite, el cual señala: la distribución de una suma de normales tiende a una normal a medida que el número de observaciones aumenta. En este punto surge un problema: las sucesivas evaluaciones a las cuales estará sometido el modelo implicarán transformaciones de las variables lo que traerá posibles repercusiones acerca de la normalidad de los residuos; por lo tanto, es importante verificar la normalidad de los residuos al principio y especialmente al final de las evaluaciones del modelo.

6.2. Evaluación Estadística: esta evaluación busca verificar si los valores de los parámetros estimados son estadísticamente significativos, porque en caso de que alguno no lo sea, no puede utilizarse para la toma de decisiones. Las pruebas estadísticas son las siguientes:

6.2.1. Prueba de significancia individual: se considera cada uno de los parámetros por separado y se sigue el siguiente procedimiento:

Por Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: $H_0: \beta_i = 0$ Hipótesis alternativa: $H_1: \beta_i \neq 0$
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	$t = \frac{\beta_i - \hat{\beta}_i}{ES(\hat{\beta}_i)}$
Regla de decisión	Si $ tc > 2$, rechazo H_0 , con un $\alpha = 0,05$, para $n > 30$. Si $\alpha > p\text{-valué}$, rechazo H_0 , con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que el parámetro (β_i) (<u>es o no es</u>) estadísticamente significativo.

En el modelo de la ecuación (1.1), haciendo la prueba de significancia individual por contraste de hipótesis sobre el parámetro β_1 , se obtienen los siguientes resultados:

Regla de decisión: $(tc = 19,4239) > 2$ ó $(\alpha = 0,05) > (p - \text{valué} = 0,000)$

Decisión: Existen evidencias para rechazar la hipótesis nula y afirmar con un nivel de significancia del 5%, que el parámetro β_1 es estadísticamente significativo.

Por Intervalos de Confianza

Un investigador necesita conocer que tan cerca está el parámetro estimado del parámetro poblacional, para ello se construyen intervalos de confianza que establecen límites alrededor del valor estimado dentro del cual esperamos encontrar el valor del verdadero parámetro, dado un cierto grado de confianza.

Se sigue el mismo procedimiento que para el contraste de hipótesis individual, con la diferencia que se debe probar la hipótesis nula sobre el siguiente intervalo de confianza:

$$\beta_i = \hat{\beta}_i \pm t_t(ES(\hat{\beta}_i))$$

Construyendo el intervalo de confianza para el parámetro β_1 para el modelo de la ecuación (1.1), se obtiene:

$$\beta_i = 0,789277 \pm 2,021(0,040634)$$

$$\beta_i = 0,789277 \pm 0,082121$$

$$0,707155 < \beta_i < 0,871398$$

Decisión: Existen evidencias para rechazar la hipótesis nula y afirmar con un nivel de significancia del 5%, que el parámetro β_1 es estadísticamente significativo.

Nota: siempre ambas pruebas por *p-valué* y por intervalo de confianza son consistentes entre sí. De no ser así, seguramente se está realizando un cálculo mal.

$ES(\hat{\beta}_i)$, es el error estándar del parámetro estimado $\hat{\beta}_i$. Para realizar estas pruebas, se utiliza en general un nivel de significancia del 5% (no se descarte un nivel de significancia del 10%) y se busca el valor del estadístico t tabulado en n-k grados de libertad, donde n es el número de observaciones y k el número de parámetros estimados.

6.2.2. Prueba de significancia conjunta: se busca comprobar si el modelo globalmente es significativo.

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: $H_0: \beta_0 = \beta_1 = 0$ Hipótesis alternativa: H_1 : al menos un parámetro difiere de cero.
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	$F = \frac{SCR * k - 1}{SCE * n - k} = \frac{(n - k) * R^2}{(k - 1) * (1 - R^2)}$
Regla de decisión	Si $ Fc > 3$, rechazo H_0 , con un $\alpha = 0,05$, para $n > 30$. Si $\alpha > p\text{-valué}$, rechazo H_0 , con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que (<u>los parámetros no son en conjunto estadísticamente significativos</u> ó <u>al menos un parámetro difiere de cero</u>).

Haciendo el contraste de hipótesis sobre el modelo (1.1), para prueba de significancia conjunta:

Regla de decisión: $(F = 377,29) > 3$ ó $(\alpha = 0,05) > (p - \text{valué} = 0,000)$

Decisión: Existen evidencias para rechazar la hipótesis nula y afirmar con un nivel de significancia del 5% que al menos un parámetro difiere de cero.

Nota: ésta evaluación se utiliza en los modelos múltiples, es decir, con más de una variable exógena, sin embargo, se explica en esta sección para no perder la secuencia del contenido de la materia. Más adelante en el segundo modelo especificado, se aplicará de nuevo.

6.2.3. Coeficiente de determinación (R^2): mide en cuando se ajusta la línea de regresión de la muestra de variables exógenas a la muestra de variable endógena, es decir, determina en que proporción la variable dependiente es explicada por las variables independientes.

$$R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT}$$

En el modelo estimado en la ecuación (1.1), el $R^2 = 0,8809$, es decir, el ingreso disponible *per cápita*, explica en un 88,09% el comportamiento del consumo *per cápita*, durante el período 1950-2002, para Venezuela.

Resultados hasta el momento

Como ya se ha demostrado, el modelo $\hat{C}_{pc} = 820,27 + 0,79Ydpc$, es económica y estadísticamente significativo y, aunque todavía falta verificar que el modelo cumpla algunos otros supuestos del MCRL, se procederá a interpretar los parámetros estimados.

$\hat{\beta}_0 = 820,27$: El valor promedio del consumo *per cápita* del venezolano durante el período 1950-2002, fue de 820,27 bs de 1984, cuando el ingreso disponible es nulo.

$\hat{\beta}_1 = 0,79$: Cuando el ingreso disponible *per cápita* del venezolano aumenta en un bolívar, el consumo *per cápita* aumenta en promedio 0,79 bs de 1984, durante el período 1950-2002.

Ejercicio:

Con base en el modelo especificado en la ecuación (7) del apartado 5.9. de relaciones funcionales, estimado en la ecuación (7.1) y expuesto a continuación: realice la evaluación económica, verifique el cumplimiento del supuesto de normalidad de los residuos y evalúe económicamente en modelo. Interprete los resultados obtenidos.

$$\begin{aligned} \text{Log}(\hat{C}_{pc}) &= -12,27 + 0,87\text{Log}(Ydpc) + 0,006\text{Tiempo} \\ &\quad t \quad \quad (-9,0486) \quad (26,439) \quad (8,9318) \\ R^2 &= 0,9649 \quad F = 686,37 \quad JB = 1,18 \quad P(JB) = 0,5534 \end{aligned}$$

6.3. Evaluación Econométrica: es esencialmente un estudio para diagnosticar la presencia y profundidad de enfermedades limitantes de la eficiencia del modelo, surgidas de la violación de algunos de los supuestos del Modelo Clásico de Regresión Lineal. Para el desarrollo de esta parte del documento en busca de un análisis más completo, necesariamente debemos partir de un modelo múltiple, para lo cual se utilizará el especificado en la ecuación (2) de la siguiente manera:

$$C = \beta_0 + \beta_1 YKd + \beta_2 YLd + \beta_3 Pob + \mu \quad (2)$$

Una vez estimado el mismo, se obtienen los siguientes resultados:

$$\begin{aligned} \hat{C} &= -5,0E^{+10} + 0,075YKd + 1,466YLd + 7350,964Pob & (2.1) \\ t &(-10,56316) (0,888187) (13,70678) (13,05268) \\ R^2 &= 0,99 \quad F = 1621,423 \quad JB = 1,44 \quad P(JB) = 0,48 \end{aligned}$$

Es de hacer notar que más allá de que $\hat{\beta}_2$ y $\hat{\beta}_3$ son estadísticamente significativos y que el estadístico Jaque-Bera permite afirmar que la inferencia es confiable, ya que, los residuos se distribuyen normal, además de que el coeficiente de determinación “parece” ser excelente y consistente con el estadístico F; los parámetros no son económicamente significativos debido a que las propensiones marginales a consumir contrastan con lo planteado en la especificación del modelo. Vale resaltar que, aunque $\hat{\beta}_1$ en efecto está entre cero y uno, no es estadísticamente significativo.

Estas inconsistencias en el modelo pueden ser producto de algunos problemas econométricos, a saber, multicolinealidad, heterocedasticidad, autocorrelación y/o errores de especificación, que a continuación se exponen.

6.3.1 Multicolinealidad: consiste en la existencia de una relación lineal perfecta, o casi perfecta, entre la totalidad, o un subgrupo, de las variables exógenas. Las causas más frecuentes de este problema son: tamaño de la muestra reducida, escasa variabilidad de las observaciones y poca calidad técnica de muestreo. Para evitar complicaciones, la base teórica del análisis de regresión múltiple asume cierta correlación entre las variables explicativas, reduciéndola para aislar su efecto individual sobre la variable endógena.

Las principales consecuencias que ocasiona el problema de multicolinealidad son: los coeficientes parciales de regresión resultan indeterminados y sus errores estándar indefinidos, ante colinealidad perfecta; los errores estándar son muy grandes con colinealidad alta pero imperfecta, causando imprecisión en la estimación de los valores poblacionales de los parámetros; se torna difícil precisar la incidencia individual de las variables exógenas sobre la endógena; las pendientes parciales estimadas pueden ser estadísticamente insignificativas con intervalos de confianza amplios que incluyen el cero, abriendo la posibilidad de que el parámetro sea igual a dicho valor; los parámetros se vuelven sensibles a las modificaciones en el tamaño de la muestra y en la estructura del modelo, reaccionando en gran medida a la exclusión de variables no significativas; finalmente, se afecta la capacidad predictiva del modelo, a menos que la multicolinealidad permanezca en las observaciones que se emplearán para predecir.

Diagnosticando la multicolinealidad

a. Diagnóstico preliminar: se basa en detectar la posible presencia de multicolinealidad por medio del coeficiente de determinación, el F y t estadísticos. En este sentido, se sospecha de multicolinealidad cuando se está en presencia de un modelo con un coeficiente de determinación alto (mayor a 0,80), prueba de hipótesis conjunta significativa (F – statistic) y pruebas de hipótesis individuales no significativas.

Nótese que en el modelo estimado en la ecuación (2.1) el parámetro $\hat{\beta}_1$ asociado a la variable YKd no es estadísticamente significativa, mientras que la prueba de significancia conjunta si lo es, al mismo tiempo que el coeficiente de determinación es mayor a 0,80; es decir, en conjunto las variables exógenas explican el comportamiento del consumo pero individualmente no, cuestión que no resulta lógica; por lo tanto se sospecha de problemas de multicolinealidad. Es necesario entonces, realizar otras pruebas que permitan constatar el grado de éste problema econométrico.

b. Correlación simple entre las regresoras: el Eviews permite calcular una matriz que contiene los coeficientes de correlación entre las variables exógenas; la regla de decisión establece que dichos coeficientes no deben ser mayores a 0,80 ya que se estaría en presencia de multicolinealidad severa. Además, estos coeficientes deben ser menores comparados con los coeficientes de correlación entre las variables exógenas y la variable endógena para tener niveles aceptables de multicolinealidad.

Para obtener la matriz de correlación, la ruta a seguir desde el menú principal es: **Quick/Group Statistic/Correlations**, y en el cuadro de diálogo emergente se debe introducir “primero la variable endógena y luego las variables exógenas” separadas por un espacio las unas de las otras. Ó desde una vista como grupo de estas variables, en el menú de trabajo: **View/Correlations/Common Sample**. Después de lo cual aparecerá la matriz de correlación como a continuación se muestra.

	Consumo	YLd	YKd	Pob
Consumo	1	0.961309546753	0.965955663882	0.902111058413
YLd	0.961309546753	1	0.94436681507	0.767391849174
YKd	0.965955663882	0.94436681507	1	0.862143037959
Pob	0.902111058413	0.767391849174	0.862143037959	1

Como se puede observar, es una matriz simétrica por debajo y por arriba de la diagonal principal que es la unidad. Los valores objetos de éste análisis son los resaltados en color amarillo que recogen la correlación parcial entre las variables exógenas. Como se pone de manifiesto en la matriz de correlación, existen evidencias para afirmar se está en presencia de problemas de multicolinealidad severa entre el ingreso disponible de los dueños del factor trabajo (YLd) y el ingreso disponible de los dueños del factor capital (YKd), ya que el coeficiente de correlación es mayor a 0,80. Igualmente sucede entre las variables Ykd y población (Pob); sin embargo no se descarta la presencia de colinealidad alta entre YLd y Pob.

c. Valor del determinante: es un valor que también permite apreciar el grado de multicolinealidad presente en el modelo. Cuando existen problemas de multicolinealidad, el valor del determinante de la matriz $X'X$ tiende a la singularidad. A medida que el valor del determinante se encuentre más cerca de cero, el problema será más severo.

Para poder realizar esta prueba es necesario hacer algunas transformaciones a las variables independientes, a saber, estandarizando y luego centralizando todas las observaciones, a través de las siguientes ecuaciones:

Para estandarizar	$X_{ki}^e = \frac{X_{ki} - \bar{X}_k}{ES(X_{ki})}$
Para centralizar	$X_{ki}^c = \frac{X_{ki}^e}{\sqrt{n - 1}}$

donde,

i = i-esima observación

k = k-esima variable independiente

X_{ki} = i-esima observación de la k-esima variable independiente

\bar{X}_k = valor promedio de la k-esima variable independiente

$ES(X_{ki})$ = error estándar de la k-esima variable independiente

n = numero de observaciones

Los pasos a seguir para calcular el valor del determinante de la matriz X'X son:

1ro. Calcular la estadística descriptiva de las variables independientes y tomar nota del valor promedio y el error estándar de cada variable.

2do. Generar las series estandarizadas y luego centralizadas (se recomienda colocar nombres nemotécnicos que permitan su fácil identificación para evitar confusiones).

3ro. Abrir y guardar las variables independientes centralizadas como grupo (nómbrelo "centralizadas").

4to. Convertir el grupo de valores centralizados en una matriz, para ello, en la barra de comandos se debe escribir: `MATRIZ M1=@CONVERT(CENTRALIZADAS)`

5to. Transponer la matriz M1, para lo cual se debe escribir en la barra de comandos lo siguiente: `MATRIZ M2=@TRANSPPOSE(M1)`

6to. Calcular el producto de la matriz transpuesta por la original ($X'X$ ó en este caso $M2M1$), a través del siguiente comando: `MATRIX PROD=M2*M1`

7mo. Finalmente calcular el valor del determinante de la matriz $X'X=M2M1=PROD$, escribiendo en la barra de comandos: `MATRIX DET=@DET(PROD)`

El resultado que en este caso arroja el valor determinante es:

R1	0.0256565648732
----	-----------------

por lo tanto, dado que el valor del determinante se acerca a cero, existen evidencias para afirmar que el problema de multicolinealidad en el modelo estimado es severo, en consonancia con lo que las pruebas anteriores han demostrado.

d. Factor de inflación de varianza: a medida que aumenta la colinealidad de los estimadores MCO, tanto varianza como su covarianza se incrementan también, a una velocidad que viene dada por:

$$FIV = \frac{1}{(1 - r^2)}$$

donde, r^2 es el coeficiente de correlación entre dos variables independientes cualesquiera, elevado al cuadrado. Para tomar la decisión acerca del grado de multicolinealidad en el modelo con base en el factor de inflación de varianza, existen las siguientes condiciones:

- 1ro. Si $FIV=0$, se concluye que la variable no está correlacionada con las demás
- 2ro. $FIV>1$, la multicolinealidad es moderada
- 3ro. $FIV>5$, la multicolinealidad es alta
- 4to. $FIV>10$, la multicolinealidad es severa

El Eviews calcula la inversa de la matriz $X'X$, es decir, $|X'X|^{-1}$, cuya diagonal principal está formada por los factores de inflación de varianza, para ello se debe escribir en la barra de comandos: `MATRIX INV=@INVERSE(PROD)`

	C1	C2	C3
R1	10.0143126994	-11.0363650303	1.82639214939
R2	-11.0363650303	16.0547938476	-5.36966954339
R3	1.82639214939	-5.36966954339	4.22837493469

Observando los valores que la diagonal principal de la matriz $|X'X|^{-1}$, resaltados en color amarillo, se puede afirmar que el problema de multicolinealidad es severo en el modelo estimado, ya que, dos factores de inflación de varianza son mayores a diez y aunque el otro no cumple esta condición está muy próximo al cinco, sobre el cual se afirma que la multicolinealidad es alta.

e. Regresiones auxiliares: esta prueba consiste en realizar regresiones de cada variable exógena en función de las restantes y, contrastar la hipótesis nula de que la variable bajo estudio es colineal con el resto de las variables exógenas, para lo cual se utiliza el estadístico F. La regla de decisión a utilizar es, si $F_c > F_t$ no se rechaza la hipótesis nula. Sin embargo es oportuno revisar el valor del coeficiente de determinación y ya que si es elevado se estará en presencia de multicolinealidad severa.

Realizando las regresiones auxiliares correspondientes al caso objeto de estudio en ésta sección del documento obtenemos:

Regresión	R ²	F _c	F _t [*]
Pob=f(YLd, YKd)	0,763528	80,72059	3,23
YLd=f(Pob, YKd)	0,900356	225,8934	3,23
YKd=f(Pob, YLd)	0,937779	376,7957	3,23

* El estadístico F_t es igual a 3,23, con 2 grados de libertad en el numerador, 40 grados de libertad en el denominador y una significancia del 5%.

Nótese como el coeficiente de determinación y el estadístico F aumentan significativamente en las dos últimas estimaciones, por lo tanto, existen evidencias para afirmar con un nivel de significancia del 5% que el problema de multicolinealidad es severo y causado por las variables YLd y YKd, con relación al resto de las variables respectivamente.

d. Mapa de Bunch: técnica que consiste en hacer regresiones que vayan incorporando las variables independientes una a una, haciendo comparaciones entre el coeficiente de determinación y la significancia individual de los parámetros, con el fin de determinar que variables están correlacionadas.

Regresión	R ²	Significancia de Parámetros
C=f(Pob)	0,813804	Significativo
C=f(YLd)	0,924116	Significativo
C=f(YKd)	0,933070	Significativo
C=f(YLd, YKd)	0,955351	Todos significativos
C=f(Pob, YLd)	0,989866	Todos significativos
C=f(Pob, YKd)	0,951789	Todos significativos
C=f(YLd, YKd, Pob)	0,990027	$\hat{\beta}_2$ no significativo

Es de hacer notar como los coeficientes de determinación de las regresiones múltiples son mayores en comparación con las regresiones simples y, que el efecto combinado de las tres variables exógenas es lo que produce el problema de multicolinealidad severa, a tal punto, que se obtiene un R² cercano a la unidad y se afecta la significancia de los parámetros.

En cuanto a la multicolinealidad: debido a la presencia de multicolinealidad en el modelo estimado, es posible que el modelo no sirva para predecir, ya que es sensible a cambios en la muestra. Los valores de las varianzas, covarianzas y errores estándar son muy grandes, alterando la contrastación de hipótesis y los intervalos de confianza son muy grandes.

Esto puede estar ocurriendo porque la muestra no es representativa para el estudio. Las posibles soluciones al problema de multicolinealidad se centran principalmente en el incremento de la base de datos, lo que resulta poco viable ya que la escasez de datos es uno de los principales problemas para el econometrista, aplicar alguna relación funcional que mejore la estimación o transformando las variables (de

nominales a reales, de agregado a *per cápita*, o diferenciar las variables). A pesar de la presencia de multicolinealidad, los estimadores MCO siguen siendo MELI.

e. Posibles soluciones

Agrandar la muestra o transformaciones económicas

En el caso que es objeto de estudio en este documento, los datos son una limitante que no permite agrandar la muestra. Por otro lado, como ya se demostró en su momento, son necesarias las correcciones de las variables por el crecimiento de los precios, sin embargo, hacer lo propio con la población no es coherente en términos económicos, por lo tanto, en este sentido, no son viables éstas posibles soluciones.

Aplicando logaritmo a las variables se obtiene:

	LOG(Consumo)	LOG(YKd)	LOG(YLd)	Pob
LOG(Consumo)	1	0.988863016446	0.987017305935	0.878681205708
LOG(YKd)	0.988863016446	1	0.9831230093	0.854404962128
LOG(YLd)	0.987017305935	0.9831230093	1	0.799602513001
Pob	0.878681205708	0.854404962128	0.799602513001	1

	LOG(Consumo)	LOG(YKd)	LOG(YLd)	LOG(Pob)
LOG(Consumo)	1	0.988863016446	0.987017305935	0.950186706074
LOG(YKd)	0.988863016446	1	0.9831230093	0.931409593362
LOG(YLd)	0.987017305935	0.9831230093	1	0.895655134899
LOG(Pob)	0.950186706074	0.931409593362	0.895655134899	1

Nótese como en la primera matriz de correlación, no se le aplicó logaritmo a la variable población, mientras que en el segundo caso si se hizo; sin embargo, en ninguno de los dos casos, la mencionada transformación logró solucionar el problema de multicolinealidad.

Aplicando diferencia a todas las variables resulta en lo siguiente:

	D(Consumo)	D(YKd)	D(YLd)	D(Pob)
D(Consumo)	1	0.220408954866	0.832914856932	-0.0770881867527
D(YKd)	0.220408954866	1	0.037707114558	-0.109369786247
D(YLd)	0.832914856932	0.037707114558	1	-0.184337466102
D(Pob)	-0.0770881867527	-0.109369786247	-0.184337466102	1

Al aplicar diferencia, se le resta el componente tendencial a las variables, lo que en este caso posiblemente cause el problema de multicolinealidad, es importante resaltar como los coeficientes de correlación disminuyen considerablemente.

Ahora bien, al estimar el modelo con el problema de multicolinealidad solucionado se obtienen la siguiente línea de regresión:

$$\Delta \hat{C} = -3,46E^{+09} + 0,133\Delta YKd + 1,267\Delta YLd + 15870,25\Delta Pob \quad (2.2)$$
$$t \quad (-0,743684) \quad (2,690271) \quad (11,24897) \quad (1,389877)$$
$$R^2 = 0,7391 \quad F = 41,33$$

Como se puede apreciar en la ecuación 2.2, el coeficiente de determinación y el estadístico F dejaron de estar inflados, al mismo tiempo los parámetros $\hat{\beta}_1$ y $\hat{\beta}_2$ son estadísticamente significativos, no obstante, $\hat{\beta}_2$ no es consistente con la teoría económica. Esta inconsistencia puede ser producto de algún otro problema econométrico en el modelo, sigamos evaluando.

6.3.2 Autocorrelación: este es un problema común de modelos de series de tiempo y se presenta cuando la perturbación asociada a cualquier observación incide, directa o inversamente, sobre la perturbación de una observación posterior indicando que son mutuamente dependientes. Las causas de la autocorrelación residual son: sesgo de especificación, fenómeno de la telaraña, comportamiento cíclico de las variables económicas y uso de modelos autorregresivos.

Al existir correlación serial de los residuos, los estimadores MCO se mantienen insesgados y consistentes, sin embargo, dejan de ser eficientes ya que, su varianza ya no es mínima, es decir, no permanecen MELI. En ese caso, las pendientes parciales de la regresión estimada ocasionan los siguientes inconvenientes: la varianza estimada es menor a la verdadera y por lo tanto se obtiene un coeficiente de determinación superior al verdadero; las pruebas t y F pierden significado, en este sentido, la prueba t se ve afectada debido a que se aceptaría la hipótesis nula cuando debería rechazarse (se comete el error de tipo I).

Es importante resaltar que la autocorrelación es un problema frecuente en datos de series de tiempo. No obstante, es factible ubicarla en series atemporales o de corte transversal, en las que la dispersión de las varianzas responde a la forma de organizar los datos y a las divergencias en las características de las distintas zonas de un país o región.

Diagnosticando la autocorrelación

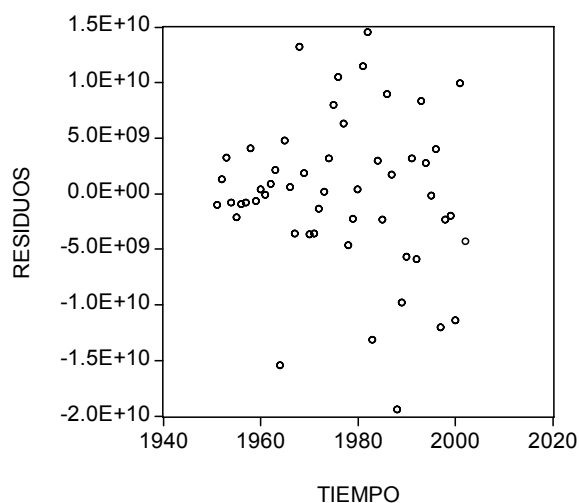
a. Prueba gráfica: consiste en realizar dos gráficos de dispersión, el primero, residuos Vs. tiempo para observarlo cuidadosamente en busca de algún patrón de comportamiento ya que, el MCRL supone la no relación entre los residuos y por lo tanto se deben comportar de manera aleatoria. El segundo gráfico a realizar es, residuos Vs. ellos mismos retardados un período, en caso de verse una correlación clara y fuerte, se podrá afirmar que existe autocorrelación de primer orden (esto significa que los residuos en el período "t" están correlacionados con sí mismos en "t-1"; en caso de segundo orden, los residuos de "t" estarán correlacionados hasta "t-2", y así sucesivamente).

Todo archivo de trabajo del Eviews tiene una variable reservada llamada "resid", en la cual se guardan los residuos de la última estimación realizada por el usuario, sin

embargo, el programa no permite hacer cálculos sobre este variable, es por ello que una vez realizada la estimación del modelo, se debe generar una serie nueva llamada residuos que se expresará en el cuadro de diálogo “*enter equation*” como: $\text{residuos}=\text{resid}$ (véase cómo generar ecuación). Otra forma más rápida de generar la serie residuos es seguir desde el menú de trabajo de la estimación la ruta: **Procs/Make a residual series**, colocar en el cuadro de diálogo emergente el nombre residuos para finalmente hacer clic en *ok*.

Luego se debe realizar el primer gráfico sobre los resultados de la ecuación 2.2, es decir, un gráfico de dispersión de los residuos en el tiempo. Como se pone de manifiesto en la gráfica 1, existe una nube de puntos de los residuos sin un claro comportamiento tendencial, cíclico y/o en definitiva un patrón, que coloque en evidencia problemas de autocorrelación en el modelo estimado.

Gráfica 1

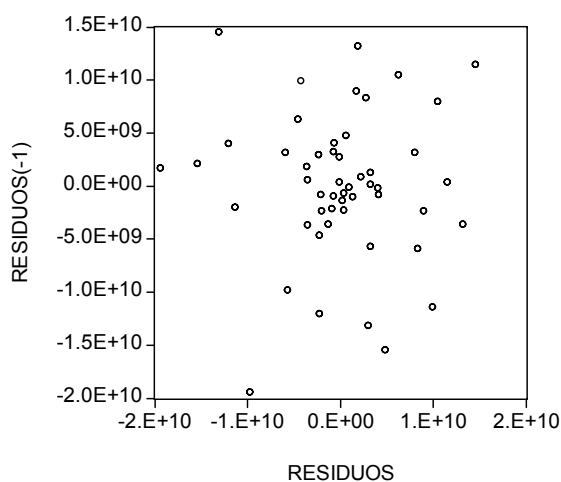


A pesar de las observaciones anteriores, se debe seguir revisando el comportamiento de los residuos para confirmar la no violación del supuesto bajo estudio. Ahora bien, como a priori no parecen haber indicios de autocorrelación, se debe suponer el problema en primer orden, en otras palabras, los residuos de hoy relacionados con los de ayer.

Basta entonces con hacer un gráfico de dispersión el cual medirá la correlación entre dos variables, que en este caso serán los residuos contra sí mismos retardados un período y, si se observa presencia de una fuerte correlación se podrá afirmar que el modelo presenta problemas de autocorrelación serial. Para ello, desde el menú principal la ruta a seguir es: **Quick/Graph/Scatter** y en el cuadro de diálogo emergente se debe escribir primero la variable se desea aparezca en el eje “x” y en segundo lugar y separado por un espacio la variable que se quiere salga en el eje “y”, que aplicado este caso será: **residuos residuos(-1)**, donde (-1) indica el retardo de la variable.

Los resultados expuestos en la gráfica 2, no muestra evidencias de autocorrelación de primer orden, en tanto que existe una nube de puntos dispersa que se traduce en una baja correlación entre las variables graficadas.

Gráfica 2



En este punto es importante resaltar que, más allá de cualquier hallazgo hecho en las pruebas gráficas, la mayoría de las veces los resultados obtenidos no son concluyentes y generan dudas al momento de decidir, por lo tanto, se deben recurrir a pruebas más formales como las que a continuación se exponen.

b. Pruebas de las Rachas o de Geary: este diagnóstico es muy sencillo debido a que no es paramétrico y tampoco necesita ningún supuesto sobre la distribución de las variables. Consiste en contrastar la hipótesis nula de no autocorrelación sobre un intervalo de confianza que viene dado por la siguiente fórmula:

$$[E(k) - 1,96 * ee_k \leq K \leq E(k) + 1,96 * ee_k]$$

Donde, $E(k)$ es la media de las rachas, ee_k es el error estándar de las rachas y, están dados por:

$$E_k = \left(\frac{2N_1N_2}{N_1 + N_2} \right) + 1$$

$$ee_k = \sqrt{\frac{2N_1N_2(2N_1N_2 - N_1 - N_2)}{(N_1 + N_2)^2(N_1 + N_2 - 1)}}$$

Para lo cual, de la serie de residuos generada por el Eviews en el “*actual,fitted,residual table*” se deben tomar: el número de signos positivos (N_1), número de signos negativos (N_2) y el número de cambios de signos o rachas (K).

Procedimiento a seguir

Planteamiento de hipótesis	Hipótesis nula: Ho: no autocorrelación Hipótesis alternativa: Hi: autocorrelación
Intervalo confianza	$[E(k) - 1,96 * ee_k \leq K \leq E(k) + 1,96 * ee_k]$
Regla de decisión	Si K no incluido en el intervalo calculado rechazo Ho
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar</u> o <u>no rechazar</u>) la hipótesis nula y afirmar que el modelo estimado (<u>presenta</u> ó <u>no presenta</u>) problemas de autocorrelación.

Los resultados obtenidos en el modelo estimado en la ecuación 2.2 son:

$$N_1 = 26; N_2 = 26; K = 28$$

$$E_k = \left(\frac{2 * 26 * 26}{26 + 26} \right) + 1 = 27$$

$$ee_k = \sqrt{\frac{2 * 26 * 26(2 * 26 * 26 - 26 - 26)}{(26 + 26)^2(26 + 26 - 1)}} = 3,57$$

Intervalo de confianza: $[27 - 1,96 * 3,57 \leq K \leq 27 + 1,96 * 3,57]$

$$[20 \leq K \leq 33,99]$$

Regla de decisión: K si se encuentra contenido en el intervalo.

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar que el modelo estimado no presenta problemas de autocorrelación.

c. Modelo Autorregresivo de Markov: esta prueba consiste en realizar una regresión donde la variable dependiente son los residuos y la independiente son los residuos con un retardo, de la siguiente manera:

$$\mu_i = \rho\mu_{t-1} + \gamma$$

Se concluirá que existe autocorrelación positiva si el coeficiente de la variable independiente es mayor que cero ($\rho > 0$). Si el coeficiente es menor que cero, se estaría en presencia de autocorrelación negativa ($\rho < 0$). Un coeficiente igual a cero ($\rho = 0$) indica que no existe autocorrelación entre los errores del modelo.

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: no autocorrelación Hipótesis alternativa: Hi: autocorrelación
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	$t = \frac{\beta_i - \hat{\beta}_i}{ES(\hat{\beta}_i)}$
Regla de decisión	Si $ tc > 2$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$. Si $\alpha > p\text{-valué}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que el modelo estimado (<u>presenta ó no presenta</u>) problemas de autocorrelación.

Aplicando el modelo autorregresivo de Markov con los residuos estimados en la ecuación 2.2, se obtiene:

$$\hat{\mu}_i = 21066892 - 0,013813\mu_{t-1}$$

$t \quad (0,021358) \quad (-0,096354)$

Regla de decisión: $(tc = -0,096354) < 2$ ó $(\alpha = 0,05) < (p - \text{valué} = 0,9236)$

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5% que el modelo estimado no presenta problemas de autocorrelación de primer orden.

d. Test Breusch-Godfrey: permite detectar la presencia de autocorrelación de mayor orden, es decir, con más de un retardo, para lo cual se determina un estadístico igual a $n * R^2$. Para realizar esta prueba la ruta a seguir desde el menú de trabajo de la estimación es: **View/Residual Tests/Serial Correlations LM Test** y en el cuadro de diálogo emergente (*Lag Specification*) se debe introducir el número de retardos a contrastar.

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: no autocorrelación Hipótesis alternativa: Hi: autocorrelación
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	$n * R^2 = Obs * R - squared \sim X^2$
Regla de decisión	Si $\alpha > p\text{-valué}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar</u> o <u>no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que el modelo estimado (<u>presenta</u> ó <u>no presenta</u>) problemas de autocorrelación.

Los resultados del estadístico de prueba están ubicados en la parte superior de la ventana, los cuales fueron calculados con base en la estimación que más abajo se presenta. Para el caso de la estimación de la ecuación 2.2 al realizar la prueba Breusch-Godfrey se obtiene:

Regla de decisión: $(\alpha = 0,05) < (p - \text{valué} = 0,913438)$

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5% que el modelo estimado no presenta problemas de autocorrelación.

e. Prueba Durbin-Watson “DW”: esta prueba es la más usada para detectar el problema de autocorrelación porque es muy sencilla y además el Eviews lo proporciona directamente en la salida de la estimación.

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: no autocorrelación Hipótesis alternativa: Hi: autocorrelación
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	DW
Regla de decisión: consiste en probar donde se encuentra DW	
Se rechaza Ho	Zona de Indecisión
0	dl
2	4-du
4	4-dl
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar</u> o <u>no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que el modelo estimado (<u>presenta</u> ó <u>no presenta</u>) problemas de autocorrelación.

* Los valores críticos dl y du se encuentran en la tabla Durbin-Watson, considerando el tamaño de la muestra y el número de variables exógenas.

Aplicando la prueba DW a la estimación de la ecuación 2.2, se obtienen los siguientes resultados:

Regla de decisión: $DW = 2,019439$; $dl = 1,421$ y $du = 1,674$

Se rechaza Ho	Zona de Indecisión	No se rechaza Ho	Zona de Indecisión	Se rechaza Ho
0	dl	du	2	4-du
4	4-dl	4	2	2,326
0	1,421	1,674	2	2,326
				2,579
				4

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5% que el modelo estimado no presenta problemas de autocorrelación.

En cuanto a la autocorrelación: a pesar del uso común de la prueba DW, presenta algunos inconvenientes: no puede ser utilizado cuando se trabaje con modelo autorregresivos, la presencia de un área de indecisión lo debilita y sólo detecta autocorrelación de primer orden.

Nótese que más allá del orden que detecte la prueba a utilizar, todos los test son consistentes entre sí, en el sentido de que todos concluyen que el modelo que es objeto de estudio no presenta problemas de autocorrelación. A pesar de esto, véase las posibles soluciones.

f. Posibles soluciones

Los métodos para corregir la autocorrelación son muy variados y su selección dependerá de si se conoce o no la estructura de la autocorrelación, es decir, si ρ es conocido. Cuando se conoce el valor de ρ se puede partir del esquema autorregresivo de primer orden para corregir satisfactoriamente el problema.

Ahora bien, es muy difícil conocer la estructura de la autocorrelación, por lo que se hace necesario recurrir a métodos que permitan su estimación. Uno de esos métodos es el procedimiento de dos etapas de Cochrane-Orcutt, la primera consiste en estimar ρ a partir de los residuales estimados, efectuando para ello el modelo de Markov ($\hat{\mu}_t = \hat{\rho}\hat{\mu}_{t-1} + \gamma$). Luego, en la segunda etapa, se utiliza la estimación de ρ para hacer la regresión de la ecuación en diferencia generalizada planteada anteriormente.

El Eviews realiza esta estimación cuando se le agrega como variable independiente un AR(1) es decir, a la línea de regresión original se le aplica un esquema autorregresivo de primer orden.

Otra opción para solucionar la violación del supuesto de no autocorrelación es estimar un modelo en diferencia, sin embargo, al tratar el problema de multicolinealidad la mencionada transformación ya se realizó. Vale la pena que el lector haga una comparación del estadístico DW antes y después de aplicar diferencia, de lo cual se observarán cambios significativos en la decisión tomada.

6.3.3 Heterocedasticidad: significa que la varianza de las perturbaciones no es constante a lo largo de las observaciones. La heterocedasticidad supone la heterogeneidad de los datos con los que se trabaja al afirmar que provienen de distribuciones de probabilidad con distinta varianza.

Algunas causas de la inconsistencia de varianza pueden ser: la capacitación, que disminuye el número de errores que se cometen en un proceso de cualquier índole donde los principiantes tienen un mayor porcentaje de errores antes de madurar en la ejecución de la actividad; el aumento del ingreso, que adjudica a los consumidores y a los entes económicos en general la potestad de escoger entre distintos patrones de gasto, de ahorro e incluso de distribución de los dividendos en el caso de las empresas con juntas accionistas. La tecnología, que facilita las labores de recolección de datos, de forma que las varianzas de las observaciones pueden disminuir a medida que se emplean equipos y procedimientos novedosos; la existencia de observaciones atípicas muy alejadas del comportamiento colectivo del resto de la muestra, además de la incorrecta especificación del modelo basada fundamentalmente en la no inclusión de variables importantes en la regresión.

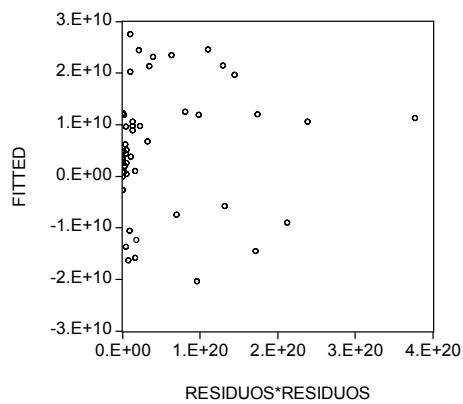
La estimación en presencia de heterocedasticidad debe realizarse cuidadosamente porque, aun cuando no elimina propiedades de insesgamiento y consistencia de los estimadores MCO, estos dejan de poseer mínima varianza perdiendo su condición de MELI. Bajo esta situación, es imposible realizar la contrastación de hipótesis de los regresores parciales, dada la gran amplitud de los intervalos de confianza, originada por la alta dispersión de las de las varianzas, aumentando la posibilidad de cometer el error de tipo II. Como última consecuencia, sobresale la baja calidad de las predicciones hechas con modelos heterocedásticos.

Diagnosticando la heterocedasticidad.

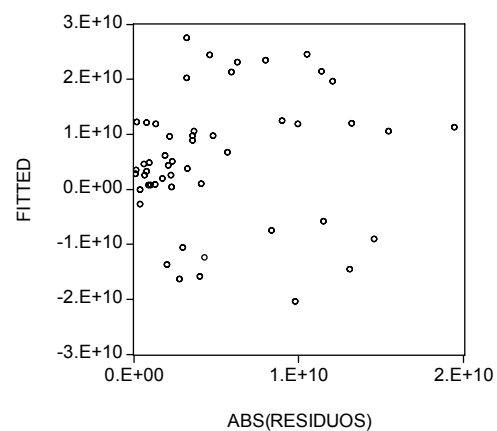
a. Prueba gráfica: esta prueba consiste en detectar el problema a través de la observación de gráficos de dispersión. Las variables a graficar son: el valor absoluto (ABS) de los residuos Vs. los valores estimados de la variable endógena para verificar la no violación del supuesto en el modelo en general. Para identificar que variable causa el

problema de heterocedasticidad de deben graficar el ABS de los residuos Vs. cada una de las variables exógenas. También se pueden utilizar los residuos elevados al cuadrado. Al observarlos se debe buscar algún patrón de comportamiento característico para concluir que el modelo presenta problemas de heterocedasticidad, sin embargo, como ya se expuso en autocorrelación, las pruebas gráficas no muestran resultados concluyentes, por lo tanto se deben recurrir a pruebas más formales.

Gráfica 3



Gráfica 4



Al observar las gráficas 3 y 4, no se evidencia de manera concluyente problemas de heterocedasticidad, ya que parece haber ausencia de algún patrón de comportamiento que lo indique.

b. Prueba de Park: Park formaliza el método gráfico, planteando que la varianza es algún tipo de función de la variable explicativa X_i . Lo que sugiere entonces, es una relación potencial, que debe ser expresada en logaritmos para aplicar el método MCO, proponiendo entonces:

$$\sigma^2_i = \sigma^2 X_i^\beta \varepsilon^{vi}$$

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: homocedasticidad Hipótesis alternativa: Hi: heterocedasticidad
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	$t = \frac{\beta_i - \hat{\beta}_i}{ES(\hat{\beta}_i)}$
Regla de decisión	Si $ tc > 2$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$. Si $\alpha > p\text{-valué}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar</u> o <u>no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que el modelo estimado (<u>presenta</u> ó <u>no presenta</u>) problemas de heterocedasticidad.

Estimado la regresión planteada por Park, con base en el modelo estimado en la ecuación 2.2 se obtiene:

$$\begin{array}{ccccccc} \text{Log}(\hat{Residuos}^2) & = & -13,16 & - & 0,12\text{Log}(\Delta YKd) & + & 0,75\text{Log}(\Delta YLd) & + & 3,38\text{Log}(\Delta Pob) \\ t & & (-0,462842) & & (-0,288092) & & (0,624592) & & (0,949335) \end{array}$$

Regla de decisión: todos los t-estadístico son menores a dos.

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5% que las variables YKd , YLd y Pob , no causan problemas de heterocedasticidad, por lo tanto el modelo estimado es homocedastico.

c. Prueba de Glejser: es muy parecida a la prueba de Park, solo que en la estimación se utiliza el valor absoluto de los residuos para calcular algunas de las siguientes regresiones:

$$|Residuos| = \beta_0 + \beta_1 X + Z$$

$$|Residuos| = \beta_0 + \beta_1 X^2 + Z$$

$$|Residuos| = \beta_0 + \beta_1 \left(\frac{1}{X}\right) + Z$$

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: homocedasticidad Hipótesis alternativa: Hi: heterocedasticidad
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	$t = \frac{\beta_i - \hat{\beta}_i}{ES(\hat{\beta}_i)}$
Regla de decisión	Si $ tc > 2$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$. Si $\alpha > p\text{-valué}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que el modelo estimado (<u>presenta</u> ó <u>no presenta</u>) problemas de heterocedasticidad.

Al estimar la regresión planteada por Glejser se obtiene:

$$ABS(\hat{Residuos}) = -3,94E^9 - 0,03\Delta YKd + 0,10\Delta YLd + 23507,87\Delta Pob$$

$$t \quad (-1,397180) \quad (-1,267222) \quad (1,487080) \quad (3,255322)$$

Regla de decisión: todos los t-estadístico son menores a dos, excepto $\hat{\beta}_3$

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5% que las variables YKd y YLd no generan problemas de heterocedasticidad, sin embargo, Pob viola el supuesto bajo estudio.

d. Prueba de White: consiste en estimar un modelo donde la variable dependiente son los residuos al cuadrado, en función de las variables independientes, sus cuadrados y sus productos cruzados. El Eviews, hace éste cálculo directamente en la salida de la estimación desde la ruta: **View/Residual Tests/White Heterocedasticity** con dos

opciones disponibles, a saber, sin términos cruzados (*no cross terms*) y con términos cruzados (*cross terms*). La selección de una u otra opción impactará sobre la robustez con la cual se tomará la decisión.

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: homocedasticidad Hipótesis alternativa: Hi: heterocedasticidad
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	$n * R^2 = Obs * R - squared \sim X^2$
Regla de decisión	Si $\alpha > p\text{-valué}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que el modelo estimado (<u>presenta</u> ó <u>no presenta</u>) problemas de heterocedasticidad.

Al aplicar el test de Withe sobre la estimación de la ecuación 2.2 resulta:

Regla de decisión: $(\alpha = 0,05) < (p - valué = 0,241949)$

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5% que el modelo estimado no presenta problemas de heterocedasticidad.

En cuanto a la heterocedasticidad: en contraste con la autocorrelación que es un problema temporal, la heterocedasticidad es espacial y por lo tanto un problema característico en datos atemporales o de corte transversal. Presenta además una dificultad al momento de su diagnostico y es que en algunas oportunidades los test no son consistentes entre sí, por lo tanto se sugiere al investigador utilice un solo método, preferiblemente el test de White, sin embargo, no se deben descartar a priori los otros.

En cuanto a las decisiones, cuando se afirma que “el modelo estimado” es o no es heterocedastico según sea el caso, esto va referido a la estimación original sobre la cual

se ésta evaluando la existencia del problema y NO a la estimación con la cual se genera el estadístico de prueba para tomar la decisión.

A pesar de las decisiones tomadas sobre el modelo de la ecuación 2.2 véase las posibles soluciones.

e. Posibles soluciones

Algunas formas de solucionar el problema de la heterocedasticidad son: aplicar logaritmo, ya que disminuye la dispersión y con ello suaviza el problema de de inflación de varianza que impacta sobre las pruebas de significancia.

Mínimos Cuadrados Consistentes con Heterocedasticidad de White, la cual realiza el Eviews si: en la salida de la estimación en el menú de trabajo se hace clic en **Estimate**, luego se debe presionar el botón **Options**, para finalmente seleccionar **Heterocedasticity Consistent Coefficient Covariance/White** y hacer clic en *ok*.

Mínimos Cuadrados Ponderados, que también desde el botón **Options** se debe seleccionar **Wheighted LS/TOLS** e introducir en **Wheight** la variable se cree genera el problema de heterocedasticidad.

En éstos dos últimos casos, en la salida de la estimación aparecerán algunas frases que indicaran el método con el cual se están estimando los parámetros.

6.3.4 Sesgo de Especificación: 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. Este procedimiento inicial que requiere selecciones entre distintas alternativas puede incurrir, sin embargo, en errores. El error de especificación es un término que cubre cualquier desviación de los supuestos del Modelo Clásico de Regresión Lineal. Cuando el modelo no aprueba algunas evaluaciones, aunque haya sido estimado por métodos no tradicionales de mayor complejidad, como el de dos etapas de Cochrane-Orcutt es necesario revisar la especificación en busca de algunas incoherencias causantes de problemas estructurales.

Algunas causas del sesgo de especificación son: inclusión de variables irrelevantes en el modelo, exclusión de variables relevantes en el modelo, forma funcional incorrecta y errores en la medición por la utilización combinada de variables expresadas en unidades físicas y unidades en distintos signos monetarios y/o años base.

La exclusión de variables relevantes es causa de inconvenientes graves, a saber: los estimadores de las variables incluidas obtenidos con el método MCO son sesgados e inconsistentes, junto a varianzas y errores incorrectos que desvirtúan las pruebas de hipótesis. Por su lado, si se incluyen variables irrelevantes, los efectos serán: los regresores parciales de las variables relevantes incluidas, así como el de las variables irrelevantes, son sesgados y consistentes, mientras que la varianza está correctamente estimada. El inconveniente más peligroso es la sobreestimación de las varianzas, pues origina intervalos de confianza más amplios de lo requerido.

Diagnosticando el sesgo de especificación.

a. Diagnostico preliminar: consiste en evaluar las pruebas t y F, para determinar la inclusión de variables irrelevantes. Haciendo el respectivo análisis sobre la ecuación 2.2 que se muestra a continuación sin problemas de multicolinealidad, autocorrelación y heterocedasticidad como se demostró con anterioridad, se puede observar que el parámetro $\hat{\beta}_3$ no es estadísticamente significativo y que la prueba de significancia conjunta es consistente con el resto de los parámetros estimados, lo que puede ser indicio de que la variable *Pob* es redundante en el modelo.

Sin embargo, una mera sospecha no puede ser determinante al momento de decidir sobre excluir o no una variable del modelo, por lo tanto se deben recurrir a pruebas más formales para ello, veamos.

b. Prueba de contribución marginal de una variable: este test tiene dos variantes, a saber, inclusión de una variable irrelevante en el modelo y omisión de una variable relevante en el modelo, las cuales descansan sobre el mismo contraste de hipótesis, por lo tanto, ya sea que se analice el problema de una variable en particular redundante o excluida, las decisiones que se toman con el test en cualquiera de sus variantes, siempre son consistentes entre sí. La ruta a seguir en el menú de trabajo de la estimación es: **View/Coefficient tests/Omitted Variables Likelihood Ratio...** (para variables relevantes omitidas) ó **View/Coefficient tests/Redundant Variables Likelihood Ratio...** (para variables irrelevantes incluidas). Una vez seleccionado el test emerge un cuadro de diálogo en el cual se debe escribir la variable se quiere sea objeto del contraste, incluso con la transformación correspondiente en caso de tenerla.

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: la contribución marginal de una variable no es significativa Hipótesis alternativa: Hi: la contribución marginal de una variable si es significativa
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	<i>F – statistic</i>
Regla de decisión	Si $\alpha > p\text{-value}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que la contribución marginal de la variable “XX” (<u>es ó no es</u>) significativa.

El caso tratado sobre la ecuación 2.2 se sospecha de la inclusión de una variable irrelevante en el modelo y de acuerdo con ello se debe aplicar el test respectivo.

Regla de decisión: $(\alpha = 0,05) < (p - \text{valué} = 0,189844)$

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5%, que la contribución marginal de la variable ΔP_{ob} no es significativa.

c. Prueba de Ramsey's Reset: los errores de especificación debidos a la formulación incorrecta de una expresión lineal para el modelo de regresión pueden ser estudiados utilizando el contraste de *Reset*. En realidad, se trata de un contraste general para detectar errores de especificación de un modelo (además de una forma funcional incorrecta, cualquier error de omisión o la presencia de correlaciones entre las variables explicativas y las perturbaciones). Ramsey demostró que la presencia en algunos casos de sesgo de especificación produce un vector de residuos de medias diferentes de cero. Desde el menú de trabajo de la estimación, la ruta a seguir es: **View/Stability Test/Ramsey RESET Test...** en el cuadro de diálogo emergente se deben escribir el número de valores estimado que corresponden a otras formas funcionales (cuadrática, cúbica...) que puedan ajustar mejor el modelo.

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: el modelo estimado esta correctamente especificado Hipótesis alternativa: Hi: el modelo estimado no esta correctamente especificado
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	<i>F - statistic</i>
Regla de decisión	Si $\alpha > p\text{-valué}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que el modelo estimado (<u>esta ó no esta</u>) correctamente especificado.

Los resultados de aplicar el test de *Reset* en el modelo de la ecuación 2.2, con dos valores estimados y un $\alpha = 0,1$ ó 10% (esto para robustecer el contraste) se exponen a continuación:

Regla de decisión: $(\alpha = 0,1) > (p - \text{valué} = 0,056974)$

Decisión: Existen evidencias para rechazar la hipótesis nula y afirmar con un nivel de significancia del 10%, que el modelo estimado no está correctamente especificado.

En cuanto al sesgo de especificación: esta evaluación consiste en detectar alguna inconsistencia en el modelo estimado no observada hasta el momento. Algunos realizan un análisis comparativo de los gráficos de los residuos con las diferentes transformaciones a las cuales ha sido sometido el modelo para verificar: si éstos empeoran con relación a la dispersión, se afirma que el modelo original está bien especificado; en caso contrario, si los residuos se han ido concentrando más a su valor medio cero, el último modelo estimado será el correctamente especificado. Sin embargo, es difícil tomar decisiones comparando los gráficos ya que la escala de medida de cada uno es diferente.

Es de hacer notar que, más allá de lo que haya dicho el test de contribución marginal de una variable, la variable *Pob* no se eliminó inmediatamente del modelo ya que, cuando se evalúa la violación del supuesto de especificación se debe ser muy cuidadoso al incluir o excluir variables en el modelo porque se podría caer de nuevo o por primera vez en sesgo de especificación. Lo recomendable es revisar en primer lugar la relación funcional o la violación de algún otro supuesto del MCRL por correcciones mal aplicadas.

d. Posibles soluciones: algunas soluciones al problema de especificación son:

Confirmar que en efecto el modelo satisfaga los supuesto de no relación lineal entre las variables exógenas, no autocorrelación y homocedasticidad, que en el caso de estudio en este documento se cumplen.

Revisar que las unidades de medida en que vienen expresadas las variables sean homogéneas, que en la ecuación 2.2 son bs de 1984. Excepto población (personas).

Comprobar la relación funcional, para ello es útil realizar gráficos de dispersión de la variable endógena con respecto a cada variable exógena o recoger algún efecto en específico aplicando logaritmo. Llevando este análisis al modelo de la ecuación 2, la idea de incluir la variable Pob es que el crecimiento poblacional impacte sobre el consumo, entonces sería pertinente obtener un coeficiente de crecimiento en β_3 para ello se le debe aplicar logaritmo al consumo. Tomando en consideración esta última transformación, es necesario entonces recoger coeficientes de elasticidad en β_1 y β_2 para no alterar el concepto de propensión marginal a consumir. Es importante resaltar que se debe aplicar logaritmo antes de diferenciar la serie, ya que en esta última transformación se pueden obtener números negativos, por lo tanto el modelo planteado sería:

$$\Delta \text{Log}(C) = \beta_0 + \beta_1 \Delta \text{Log}(YKd) + \beta_2 \Delta \text{Log}(YLd) + \beta_3 \Delta Pob + \mu$$

Que una vez estimado se obtiene:

$$\begin{aligned} \Delta \text{Log}(\hat{C}) &= 0,033 + 0,115 \Delta \text{Log}(YKd) + 0,601 \Delta \text{Log}(YLd) - 4,13E^{-8} \Delta Pob & (2.3) \\ t &(0,966770) \quad (2,087740) & (6,984069) \quad (-0,479321) \\ R^2 &= 0,6274 \quad F = 26,94686 \end{aligned}$$

Nótese que el modelo mejora en términos de significancia económica, ya que se obtienen en efecto valores entre cero y la unidad de acuerdo al concepto de propensión marginal a consumir en $\hat{\beta}_1$ y $\hat{\beta}_2$. Al mismo tiempo el parámetro $\hat{\beta}_3$ no es económicamente significativo, en el sentido de que el crecimiento poblacional favorece el consumo y el signo esperado debería ser positivo, pero no es así; esto puede ser debido a que al diferenciar la serie se le resta el componente tendencial y ya no puede considerar el efecto crecimiento. Sumado a ello, el test de contribución marginal de la variable Pob indicó que es redundante en el modelo, por lo que es pertinente excluirla. En definitiva el modelo estimado quedaría de la siguiente manera:

$$\begin{aligned} \Delta \text{Log}(\hat{C}) &= 0,017 + 0,1219 \Delta \text{Log}(YKd) + 0,6163 \Delta \text{Log}(YLd) & (2.4) \\ t &(2,465157) \quad (2,308243) & (7,776501) \\ R^2 &= 0,6256 \quad F = 40,94911 \end{aligned}$$

7. Quinta Etapa del Proceso Econométrico:

Capacidad Predictiva del Modelo: Función de Consumo

Una vez obtenido el modelo final, que en este caso fue el estimado en la ecuación 2.4 y que además debe cumplir con todos los supuestos del modelo clásico de regresión lineal¹, sólo resta evaluar la capacidad predictiva del modelo a través de los siguientes test.

7.1 Estabilidad paramétrica: consiste en probar que los parámetros son de valor constante para todo el periodo muestral, es decir, su magnitud no cambia aún y se divide en sub periodos. Para ello, existe el test de estabilidad paramétrica de Chow que pretende verificar la existencia de un cambio estructural, que puede ser dos interceptos ó, pendientes diferentes en el modelo estimado. La ruta a seguir desde el menú de trabajo de la estimación es: **View/Stability Tests/Chow Beakpoint Test...** después de los cual emergerá un cuadro de diálogo en cual se debe introducir la fecha donde se piensa ocurrió el cambio estructural. Éste último hecho, el de introducir la fecha, es precisamente una de las mayores debilidades del test de Chow, ya que necesariamente se debe conocer el contexto histórico en que se desarrolla el fenómeno estudiado, cuál ha sido el comportamiento de la serie temporal y cómo aquél de manera cuantitativa impacto sobre éste.

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: los parámetros son estables Hipótesis alternativa: Hi: los parámetros no son estables
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	<i>F – statistic</i>
Regla de decisión	Si $\alpha > p\text{-valué}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.

¹ El modelo estimado en la ecuación 2.4 a pesar de que cumple todos los supuestos del MCRL, no satisface el de normalidad de los residuos como consecuencia de la última transformación a la cual estuvo sujeto, por lo tanto no se puede confiar en las predicciones hechas con el mismo. Sin embargo, a los fines de este documento, omitiremos esta inconsistencia para darle continuidad al contenido.

Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que los parámetros estimados (<u>son ó no son</u>) estables.
-----------------------------	--

Para determinar el año del cambio estructural en el modelo de la ecuación 2.4 se debe considerar la gráfica 1, de la sección 4.4 de este documento, en la cual se expone el comportamiento del consumo durante el período bajo estudio. Como se puede observar, desde 1950 el consumo ha tenido un comportamiento creciente y sostenido, sin embargo, a partir de 1978, se produce un significativo cambio estructural donde se estabiliza la variable sin un aparente componente tendencial, de allí que el año que se considere al aplicar el test del Chow sea 1978.

Regla de decisión: $(\alpha = 0,05) < (p - \text{valué} = 0,279517)$

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5%, que los parámetros estimados son estables.

Es de hacer notar, la contradicción que existe entre el cambio estructural diagnosticado por observación gráfica con relación a lo que el test de Chow indica. Esto ocurre ya que, en el modelo estimado en la ecuación 2.4 no se está modelando el consumo en su estado natural, sino como la diferencia del mismo (ΔC), transformación ésta que le resta el componente tendencial a la serie y por lo tanto suprime el cambio estructural observado. Para efectos de demostración se puede graficar la serie en su estado natural y luego diferenciada.

7.2 Predicción: permite determinar que tan bueno es el modelo para predecir. Para usarla es necesario que el modelo satisfaga todos los supuestos del modelo clásico de regresión lineal y la normalidad de los residuos. Para evaluar la capacidad predictiva del modelo se tienen dos opciones:

a. Prueba de Predicción de Chow: para el cual también debe conocerse al año que se supone ocurrió el cambio estructural y a partir de allí se harán predicciones, sobre

las cuales se determinará su consistencia en el ajuste, generando un estadístico para contratación. Desde el menú de trabajo la ruta a seguir es: **View/Stability Tests/Chow Forecast Test...**

Contraste de Hipótesis

Planteamiento de hipótesis	Hipótesis nula: Ho: los parámetros son consistentes Hipótesis alternativa: Hi: los parámetros no son consistentes
Nivel de significancia	$\alpha = 0,05$ ó 5%
Estadístico de prueba	<i>F – statistic</i>
Regla de decisión	Si $\alpha > p\text{-valué}$, rechazo Ho, con un $\alpha = 0,05$, para $n > 30$.
Decisión (fórmula genérica)	Existen evidencias para (<u>rechazar o no rechazar</u>) la hipótesis nula y afirmar con un nivel de significancia del 5%, que los parámetros estimados (<u>son ó no son</u>) consistentes.

Los resultados obtenidos luego de aplicar el test de predicción de Chow al modelo estimado en la ecuación 2.4 y considerar 1978 como año de cambio estructural son:

Regla de decisión: $(\alpha = 0,05) < (p - \text{valué} = 0,997276)$

Decisión: Existen evidencias para no rechazar la hipótesis nula y afirmar con un nivel de significancia del 5%, que los parámetros estimados son consistentes.

b. Forecast: Eviews contiene en el menú de trabajo de la estimación la opción de predicción (*Forecast*) y de los resultados obtenidos allí se puede hacer inferencia acerca de la capacidad predictiva del modelo estimado. Las opciones que se incluyen en este cuadro relativo al cálculo de las predicciones son las siguientes:

1ro. *Forecast name:* introducción de un nombre para la serie de las predicciones.

2do. *S.E. (optional)*: la posibilidad de salvar los valores de los errores estándar de la predicción si en esta casilla escribimos un nombre.

3ro. *Method*: elección del método de cálculo de las predicciones entre: dinámico (*Dynamic*), estático (*Static*) o estructural (*Structural*). El método dinámico obtiene las predicciones que se calculan a partir de la primera observación del período de predicción y utilizando valores predichos en el caso de que existan variables retardadas de la endógena. El método estático calcula las predicciones una a una a partir de la observación anterior y utilizando valores actuales.

4to. *Sample range for forecast*: tamaño del período de predicción.

5to. *Output*: presentación de los resultados de diversas formas: representación gráfica (*Do graph*), una tabla con medidas que analizan la capacidad predictiva (*Forecast evaluation*) o ambas cosas.

Los estadísticos que computa el Eviews son:

1ro. Raíz del error cuadrático (*Root Mean Squared Error*)

2do. Error absoluto medio (*Mean Absolute Error*)

3ro. Error absoluto medio del porcentaje de error (*Mean Abs. Percent Error*)

4to. Coeficiente de desigualdad de Theil (*Theil Inequality Coefficient*)

Todos los estadísticos descritos hasta ahora indican una mejor capacidad predictiva del modelo cuanto más cercanos a cero estén, lo que permite comparar un determinado modelo con otros alternativos.

La descomposición del error cuadrático medio de predicción genera un conjunto de componentes que también se utilizan para valorar los resultados predictivos de un modelo. El cociente entre cada uno de los componentes en la suma total se denomina:

1ro. Proporción del sesgo (*Bias Proportion*)

2do. Proporción de la varianza (*Variance Proportion*)

3ro. Proporción de la covarianza (*Covariance Proportion*)

Cada una de estas proporciones varía entre cero y uno, siendo su suma la unidad. Los dos primeros miden, respectivamente, las diferencias entre la media y la varianza de la serie predicha y las de la serie observada. Por lo tanto, lo deseable es que su valor sea pequeño. La última proporción mide la parte residual o no sistemática de los errores de predicción, en donde debería recaer la mayor parte del error total cometido.

Regla de decisión: si la raíz del error cuadrático medio, el error absoluto medio y el error absoluto medio del porcentaje de error, son menores a tres, el modelo estimado tiene buena capacidad predictiva.

Decisión: Existen evidencias para afirmar que el modelo (tienen o no tiene) buena capacidad predictiva.

Aplicando este método al modelo estimado en la ecuación 2.4 se tiene:

Regla de decisión: *Mean Abs. Percent Error* = 8,56 > 3

Decisión: Existen evidencias para afirmar que el modelo no tiene buena capacidad predictiva.

En cuanto a la predicción: con esta última etapa queda finalizado el proceso econométrico como técnica de cuantificación, ahora bien, que un modelo tenga o no buena capacidad predictiva, sólo impacta en cuanto a la toma de decisiones con base a objetivos específicos planteados. Así por ejemplo, en el caso del modelo 2.4 que no tiene buena capacidad predictiva, lógicamente, no se podrán tomar decisiones de política económica con relación al fenómeno considerado, sin embargo, es útil para hacer una evaluación post mortem.

Ejercicio

Con base en el modelo estimado en la ecuación 1.1, determine si existe algún cambio estructural y evalúe la estabilidad paramétrica y predicción del modelo.

TOPICOS DE INTERES

1. Variables Cualitativas

Las variables consideradas hasta el momento para los modelos de regresión son de tipo cuantitativo; sin embargo, existen variables de tipo cualitativo que pueden ser relevantes para explicar el comportamiento de una variable endógena.

Éstas variables se conocen como: ficticias, artificiales, categóricas, cualitativas, dicotómicas o dummy.

1.1 Características

1ro. Las variables dicotómicas toman dos valores asignados de manera arbitraria, generalmente, 1 si la observación presenta la característica y 0 en caso contrario. Por ejemplo:

$$D_{1i} = \begin{cases} 1 & \text{si } i \text{ tiene hijos} \\ 0 & \text{si } i \text{ no tiene hijos} \end{cases}$$

2do. La regla general es, introducir en un modelo tantas variables dicotómicas como categorías tenga el factor cualitativo menos una; es decir, para “**m**” categorías, “**m-1**” variables dicotómicas. En caso de introducir “**m**” variables dicotómicas se caería en la “**Trampa de la Variable Dicotómica**” (colinealidad perfecta); trampa que se obvia si no se incluye el intercepto en el modelo.

3ro. La categoría a la cual no se le asigna una variable dicotómica, se le conoce como base, de comparación, de control o referencia. El intercepto representa el valor promedio de la categoría control.

4to. El uso de variables dicotómicas en los modelos de regresión pueden ser de dos tipos:

a. Como una variable dependiente → Modelos de Respuesta Cualitativa → Método de estimación: LOGIT y PROGIT.

b. Como variable independiente → Dos casos:

{ Todas las variables independientes son cualitativas → Modelos de Análisis de Varianza (ANOVA).
{ Variables independientes cuantitativas y cualitativas → Modelos de Análisis de Covarianzas (ANCOVA).

1.2 Modelos ANOVA

Considérese el siguiente modelo: Ingreso = f(Nivel Educativo)

Nótese que “Ingreso” es la variable endógena y cuantitativa, mientras que “Nivel Educativo” es la variable exógena y cualitativa expresada en cuatro categorías, las cuales son: sin estudios, primaria, secundaria, universitario. Obteniéndose a partir de éstas las siguientes variables dicotómicas:

$$\begin{aligned} \text{Sinest} & \begin{cases} 1 \text{ si no tiene estudios} \\ 0 \text{ otro caso} \end{cases} \\ \text{Prim} & \begin{cases} 1 \text{ si tiene primaria} \\ 0 \text{ otro caso} \end{cases} \\ \text{Secun} & \begin{cases} 1 \text{ si tiene secundaria} \\ 0 \text{ otro caso} \end{cases} \\ \text{Univer} & \begin{cases} 1 \text{ si tiene estudios universitarios} \\ 0 \text{ otro caso} \end{cases} \end{aligned}$$

El modelo planteado es:

$$\text{Ingreso} = \beta_1 \text{Sinestu} + \beta_2 \text{Prim} + \beta_3 \text{Secun} + \beta_4 \text{Unver} + \mu$$

Vale la resaltar que, el modelo planteado “no tiene intercepto” para evitar caer en la trampa de la variable dicotómica.

Dependent Variable: INGRESO				
Method: Least Squares				
Sample: 1 500				
Included observations: 500				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
SINEST	71884.43	6299.592	11.41096	0.0000
PRIM	85618.65	5750.715	14.88835	0.0000
SECUN	106303.8	13899.73	7.647902	0.0000
UNIVER	197746.6	11553.59	17.11560	0.0000
R-squared	0.161953	Mean dependent var	94443.14	
Adjusted R-squared	0.156884	S.D. dependent var	93315.68	
S.E. of regression	85683.72	Akaike info criterion	25.56268	
Sum squared resid	3.64E+12	Schwarz criterion	25.59640	
Log likelihood	-6386.670	Durbin-Watson stat	1.934322	

$$\widehat{\text{Ingreso}} = 71884.43 \text{Sinest} + 85618.65 \text{Prim} + 106303.8 \text{Secun} + 197746.6 \text{Unver}$$

Interpretación

1ro. Si la persona no tiene estudios, “Sinest=1”, “Prim=0”, “Secun=0” y “Univer=0”. De acuerdo con lo anterior, el modelo resulta de la siguiente manera:

$$\widehat{\text{Ingreso}} = 71884.43(1) + 85618.65(0) + 106303.8(0) + 197746.6(0)$$

$$\widehat{\text{Ingreso}} = 71884.43$$

$\hat{\beta}_1 = 71884.43$: El valor promedio del ingreso para las personas sin estudios es de 71884.43 bolívares.

2do. Si la persona tiene primaria, “Sinest=0”, “Prim=1”, “Secun=0” y “Univer=0”. Resolviendo el modelo obtenemos:

$$\widehat{\text{Ingreso}} = 71884.43(0) + 85618.65(1) + 106303.8(0) + 197746.6(0)$$

$$\widehat{\text{Ingreso}} = 85618.65$$

$\hat{\beta}_2 = 85618.65$: El valor promedio del ingreso para las personas con primaria es de 85618.65 bolívares.

3ro. Si la persona tiene secundaria, “Sinest=0”, “Prim=0”, “Secun=1” y “Univer=0”. Por lo tanto se tiene que:

$$\widehat{\text{Ingreso}} = 71884.43(0) + 85618.65(0) + 106303.8(1) + 197746.6(0)$$

$$\widehat{\text{Ingreso}} = 106303.8$$

$\hat{\beta}_3 = 106303.8$: El valor promedio del ingreso para las personas con secundaria es de 106303.8 bolívares.

4to. Si la persona tiene estudios universitarios, “Sinest=0”, “Prim=0”, “Secun=0” y “Univer=1”. Por lo tanto se tiene que:

$$\widehat{\text{Ingreso}} = 71884.43(0) + 85618.65(0) + 106303.8(0) + 197746.6(1)$$

$$\widehat{\text{Ingreso}} = 197746.6$$

$\hat{\beta}_4 = 197746.6$: El valor promedio del ingreso para las personas con estudios universitarios es de 197746.6 bolívares.

Sin embargo, lo usual es plantear modelos econométricos con intercepto y, en los Modelos ANOVA también se puede hacer, no obstante, para evitar caer en la trampa de la variable dicotómica, se debe eliminar una variable dicotómica de las definidas con anterioridad, la cual será tomada como categoría control. En este caso, se tomará como categoría control la variable “sin estudios” y así se dará cumplimiento a la característica número dos de las variables dicotómicas, es decir, para “m” categorías (m=4), “m-1” variables dicotómicas (m-1=3).

El modelo entonces quedará replanteado de la siguiente manera:

$$\text{Ingreso} = \beta_0 + \beta_1 \text{Prim} + \beta_2 \text{Secun} + \beta_3 \text{Unver} + U$$

Recuerde que el intercepto representa el valor promedio de la categoría control.

Dependent Variable: INGRESO				
Method: Least Squares				
Sample: 1 500				
Included observations: 500				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
PRIMARIA	13734.23	8529.688	1.610167	0.1080
SECUNDARIA	34419.39	15260.65	2.255434	0.0245
UNIVERSITARIO	125862.2	13159.42	9.564418	0.0000
C	71884.43	6299.592	11.41096	0.0000
R-squared	0.161953	Mean dependent var		94443.14
Adjusted R-squared	0.156884	S.D. dependent var		93315.68
S.E. of regression	85683.72	Akaike info criterion		25.56268
Sum squared resid	3.64E+12	Schwarz criterion		25.59640
Log likelihood	-6386.670	F-statistic		31.95070
Durbin-Watson stat	1.934322	Prob(F-statistic)		0.000000

$$\widehat{\text{Ingreso}} = 71884.43 + 13734.23\text{Prim} + 34419.39\text{Secun} + 125862.2\text{Unver}$$

Interpretación

1ro. Si la persona no tiene estudios, “Prim=0”, “Secun=0” y “Univer=0”. De acuerdo con lo anterior, el modelo resulta de la siguiente manera:

$$\widehat{\text{Ingreso}} = 71884.43 + 13734.23(0) + 34419.39(0) + 125862.2(0)$$

$$\widehat{\text{Ingreso}} = 71884.43$$

$\hat{\beta}_0 = 71884.43$: El valor promedio del ingreso para las personas sin estudios es de 71884.43 bolívares.

2do. Si la persona tiene primaria, “Prim=1”, “Secun=0” y “Univer=0”. Resolviendo el modelo obtenemos:

$$\widehat{\text{Ingreso}} = 71884.43 + 13734.23(1) + 34419.39(0) + 125862.2(0)$$

$$\widehat{\text{Ingreso}} = 85618.65$$

$\hat{\alpha}_1 = (\hat{\beta}_0 + \hat{\beta}_1) = 85618.65$: El valor promedio del ingreso para las personas con primaria es de 85618.65 bolívares.

$\hat{\beta}_1 = 13734.23$: El valor promedio del ingreso para las personas con primaria es de 13734.23 bolívares mayor al de las personas sin estudios.

3ro. Si la persona tiene secundaria, “Prim=0”, “Secun=1” y “Univer=0”. Por lo tanto se tiene que:

$$\widehat{\text{Ingreso}} = 71884.43 + 13734.23(0) + 34419.39(1) + 125862.2(0)$$

$$\widehat{\text{Ingreso}} = 106303.8$$

$\hat{\alpha}_2 = (\hat{\beta}_0 + \hat{\beta}_2) = 106303.8$: El valor promedio del ingreso para las personas con secundaria es de 106303.8 bolívares.

$\hat{\beta}_2 = 34419.39$: El valor promedio del ingreso para las personas con secundaria es de 34419.39 bolívares mayor al de las personas sin estudios.

4to. Si la persona tiene estudios universitarios, “Prim=0”, “Secun=0” y “Univer=1”. Por lo tanto se tiene que:

$$\widehat{\text{Ingreso}} = 71884.43 + 13734.23(0) + 34419.39(0) + 125862.2(1)$$

$$\widehat{\text{Ingreso}} = 197746.6$$

$\hat{\alpha}_3 = (\hat{\beta}_0 + \hat{\beta}_3) = 197746.6$: El valor promedio del ingreso para las personas con estudios universitarios es de 197746.6 bolívares.

$\hat{\beta}_3 = 125862.2$: El valor promedio del ingreso para las personas con estudios universitarios es de 125862.2 bolívares mayor al de las personas sin estudios.

1.3 Modelos ANCOVA

Considérese el siguiente modelo: $Gasto = f(\text{Ingreso}, \text{Lugar de residencia})$

Nótese que las variables “Gasto” e “Ingreso”, son cuantitativas, la primera endógena y la segunda exógena; al mismo tiempo que la variable “Lugar de residencia”, es cualitativa y contiene dos categorías: si la persona reside en zona urbana y si la persona reside en zona rural. A partir de éstas y considerando la trampa de la variable dicotómica y que se estimará un modelo con intercepto, se obtiene la siguiente variable dicotómica:

$$\text{Urbano} \begin{cases} 1 & \text{si la persona reside en zona urbana} \\ 0 & \text{otro caso} \end{cases}$$

El modelo planteado es:

$$Gasto = \beta_0 + \beta_1 \text{Ingreso} + \beta_2 \text{Urbano} + \mu$$

donde,

β_0 : representa el valor promedio del gasto para las personas que residen en zona rural cuando el ingreso es nulo.

β_1 : representa el valor promedio de la variación en el gasto, dado un incremento de una unidad en el ingreso para las dos categorías.

β_2 : representa la diferencia entre el valor promedio del gasto para los residentes en la zona rural y urbana cuando el ingreso es nulo.

Dependent Variable: GASTO				
Method: Least Squares				
Sample: 1 500				
Included observations: 500				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
INGRESO	0.045203	0.002805	16.11743	0.0000
URBANO	19207.19	6775.803	2.834673	0.0048
C	7312.043	6788.187	1.077172	0.2819
R-squared	0.381856	Mean dependent var	94443.14	
Adjusted R-squared	0.379368	S.D. dependent var	93315.68	
S.E. of regression	73514.28	Akaike info criterion	25.25433	
Sum squared resid	2.69E+12	Schwarz criterion	25.27962	
Log likelihood	-6310.582	F-statistic	153.5096	
Durbin-Watson stat	2.136714	Prob(F-statistic)	0.000000	

$$\widehat{Gasto} = 7312.043 + 0.045203Ingreso + 19207.19Urbano$$

Interpretación

$\hat{\beta}_0 = 7312.043$: El valor promedio del gasto para las personas que viven en zona rural es de 7312.03 bolívares, cuando el ingreso es nulo.

$\hat{\beta}_1 = 0.045203$: Por cada aumento de un bolívar en el ingreso, el gasto aumenta en promedio 0.045203 bolívares para las dos categorías.

$\hat{\beta}_3 = 19207.19$: El valor promedio del gasto para las personas que residen en zona urbana es de 19207.19 bolívares mayor al de las personas que viven en zona rural, cuando el ingreso es nulo ó, $\hat{\alpha}_3 = (\hat{\beta}_0 + \hat{\beta}_3) = 26519.23$: El valor promedio del gasto para las personas que residen en zona urbana es de 26519.23 bolívares cuando el ingreso es nulo.

1.3.1 Efecto Intercepto

Modelo: Gasto = f(Ingreso, Lugar de residencia)

$$Gasto = \beta_0 + \beta_1 Ingreso + \beta_2 Urbano + \mu$$

Al incluir en el modelo la variable dicotómica “Urbano” se obtiene un efecto desplazamiento en el intercepto que se debe a la zona en que reside la persona, por su lado, el efecto del ingreso sobre el gasto para las dos categorías es constante, lo que se traduce en la misma pendiente, de la siguiente manera:

1ro. Modelo de gasto para las personas que residen en zona rural:

$$\widehat{Gasto} = \hat{\beta}_0 + \hat{\beta}_1 Ingreso$$

2do. Modelo de gasto para las personas que residen en zona urbana:

$$\widehat{Gasto} = (\hat{\beta}_0 + \hat{\beta}_2) + \hat{\beta}_1 Ingreso$$

1.3.2 Efecto Pendiente

Modelo: Gasto = f(Ingreso, Lugar de residencia)

Considérese ahora la interacción de variables

$$Gasto = \beta_0 + \beta_1 Ingreso + \beta_2 Urbano + \beta_3 Ingreso * Urbano + \mu$$

Al incluir en el modelo la variable dicotómica “Urbano” y la interacción de la misma con el ingreso, se obtiene además del efecto intercepto, un efecto pendiente que en ambos casos que se debe a la zona en que reside la persona, de la siguiente manera:

1ro. Modelo de gasto para las personas que residen en zona rural:

$$\widehat{Gasto} = \hat{\beta}_0 + \hat{\beta}_1 Ingreso$$

2do. Modelo de gasto para las personas que residen en zona urbana:

$$\widehat{Gasto} = (\hat{\beta}_0 + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_3) Ingreso$$

Dependent Variable: GASTO				
Method: Least Squares				
Sample: 1 500				
Included observations: 500				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
INGRESO	0.040334	0.004472	9.020111	0.0000
URBANO	3518.340	13110.23	0.268366	0.7885
INGRESO*URBANO	0.008018	0.005738	1.397372	0.1629
C	1115.768	9075.601	0.122942	0.9022
R-squared	0.384279	Mean dependent var	94443.14	
Adjusted R-squared	0.380555	S.D. dependent var	93315.68	
S.E. of regression	73443.92	Akaike info criterion	25.25440	
Sum squared resid	2.68E+12	Schwarz criterion	25.28812	
Log likelihood	-6309.600	F-statistic	103.1868	
Durbin-Watson stat	2.141086	Prob(F-statistic)	0.000000	

$$\widehat{Gasto} = 1115.768 + 0.040334 Ingreso + 3518.340 Urbano + 0.008018 Ingreso * Urbano$$

Interpretación

$\hat{\beta}_0 = 1115.768$: El valor promedio del gasto para las personas que viven en zona rural es de 1115.768 bolívares, cuando el ingreso es nulo.

$\hat{\beta}_1 = 0.040334$: Por cada aumento de un bolívar en el ingreso, el gasto aumenta en promedio 0.040334 bolívares para las personas que residen en zona rural.

$\hat{\beta}_2 = 3518.340$: El valor promedio del gasto para las personas que residen en zona urbana es de 3518.340 bolívares mayor al de las personas que residen en zona rural cuando el ingreso es nulo ó, $\hat{\alpha}_2 = (\hat{\beta}_0 + \hat{\beta}_2) = 4634.11$: El valor promedio del gasto para las personas que residen en zona urbana es de 4634.11 bolívares cuando el ingreso es nulo.

$\hat{\beta}_3 = 0.008018$: Por cada aumento de un bolívar en el ingreso, el gasto aumenta en promedio 0.008018 bolívares más para las personas que residen en zona urbana con respecto a las que residen en zona rural ó, $\hat{\alpha}_3 = (\hat{\beta}_1 + \hat{\beta}_3) = 0.048352$: Por cada aumento de un bolívar en el ingreso, el gasto aumenta en promedio 0.048352 bolívares para las personas que residen en zona urbana.

2. Cointegración

2.1 Procesos Estocásticos

Un proceso estocástico es el conjunto de variables aleatorias que corresponden a momentos sucesivos del tiempo. Cuando se observe una serie concreta tomando valores $X_1, X_2, X_3, \dots, X_t$, diremos que representa una muestra del proceso o una realización del mismo.

De la misma forma en que utilizamos los datos muestrales para hacer inferencias respecto a la población, en las series de tiempo se emplea la realización para llevar a cabo inferencias respecto al proceso estocástico subyacente. La distinción entre proceso estocástico y su realización es semejante a la diferencia entre población y muestra en datos transversales.

2.2 Procesos Estocásticos Estacionarios

Al hablar de procesos estacionarios se puede hacer en sentido amplio o en sentido estricto. En sentido amplio, una serie de tiempo es estacionaria si su media, su varianza y su autocovarianza permanecen iguales sin importar en momento en el cual se midan; es decir, son invariantes respecto al tiempo. Por su lado, una serie de tiempo es estrictamente estacionaria si todos los momentos de su distribución de probabilidad (y no sólo los dos primeros, es decir, la media y la varianza) son invariantes respecto al tiempo.

Un tipo especial de proceso estocástico es el **proceso puramente aleatorio** o de **ruido blanco**, el cual tiene una media igual a cero, una varianza constante σ^2 y no esta serialmente correlacionada. Recuérdese que el término de error μ_t que está presente en el modelo clásico de regresión lineal, se supuso que era un proceso ruido blanco denotado por $\mu_t \sim IIDN(0, \sigma^2)$; es decir, μ_t está distribuida de manera independiente e idéntica que la distribución normal con media cero y varianza constante.

2.3 Procesos Estocásticos No Estacionarios

En sentido amplio, una serie de tiempo no estacionaria tendrá una media que varía con el tiempo o una varianza que cambia con el tiempo, o ambas. En general, tres son los motivos por los cuales no suele producirse estacionariedad: la serie presenta tendencia, la varianza no es constante o la serie presenta variaciones no estables en intervalos regulares de tiempo.

Si la no estacionariedad es debida a la presencia de tendencia, si se diferencia la serie en el tiempo, normalmente se consigue que la serie resulte sea estacionaria:

$$Z_t = \Delta X_t = X_t - X_{t-1}$$

Si la serie presenta tendencia lineal, una sola diferencia será suficiente para convertirla en estacionaria. Si la serie presenta tendencia cuadrática, será necesaria una segunda diferencia. El número de diferencias se conoce como orden de integración.

Si X_t es no estacionaria, pero $Z_t = \Delta^d X_t$ es estacionaria $\rightarrow X_t \sim I(d)$.

Por otra parte, si observamos que la varianza de la serie no es constante a lo largo del tiempo, es útil realizar una transformación logarítmica de la serie con anterioridad a tomar diferencias. Si tras dicha transformación, solamente hemos corregido la no estacionariedad en varianza, entonces será necesario tomar diferencias sobre la transformación logarítmica.

Las series que no logran hacerse estacionarias a través de la diferenciación se llaman no homogéneas o explosivas.

2.4 Importancia de la Estacionariedad

1ro. Fenómeno de la **regresión espuria**. El análisis de regresión con variables no estacionarias, puede dar resultados sin sentido (R^2 alto, t significativo pero DW malo).

2do. El análisis de regresión basado en información de series de tiempo supone implícitamente que las series de tiempo en la cual se basa son estacionarias. Las pruebas clásicas t y F entre otras, están basadas en este supuesto.

3ro. Granger y Newbold (1974) determinación que el 75% de las veces se rechaza erróneamente la hipótesis nula de no significación, encontrando relaciones significativas donde no las hay.

4to. Los estimadores mínimos cuadráticos, no son consistentes, pues no convergen al verdadero valor, ante la presencia de series de tiempo no estacionarias.

5to. En la práctica, la mayoría de las series económicas son no estacionarias.

2.5 Proceso Estocástico de Raíz Unitaria

Este proceso se describe como:

$$Y_t = \rho Y_{t-1} + \mu_t \quad -1 \leq \rho \leq 1$$

Si $\rho = 1$, se tiene lo que se conoce como problema de raíz unitaria, es decir, se enfrenta una situación de no estacionariedad. Efecto permanente en el tiempo.

Si $|\rho| < 1$, se puede demostrar que la serie de tiempo Y_t es estacionaria. Efecto que se desvanece en el tiempo.

Diagnosticando la Estacionariedad

2.5.1 Prueba Gráfica

Observe si la variable presenta tendencia o fluctuaciones que indique que su varianza no es constante.

2.5.2 Prueba de Raíz Unitaria

a. Test Dickey-Fuller (DF): el punto de inicio es el proceso estocástico de raíz unitaria:

$$Y_t = \rho Y_{t-1} + \mu_t \quad -1 \leq \rho \leq 1$$

donde μ_t es un término de error con ruido blanco. La idea general detrás de la prueba de raíz unitaria, consiste en estimar la regresión de Y_t sobre su valor rezagado (de un período) Y_{t-1} y se averigua si la ρ estimada es estadísticamente igual a uno.

Por razones teóricas, se manipula la ecuación inicial restando Y_{t-1} en ambos lados, obteniéndose:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} + \mu_t$$

$$Y_t - Y_{t-1} = (\rho - 1)Y_{t-1} + \mu_t$$

$$\Delta Y_t = \delta Y_{t-1} + \mu_t$$

donde $\delta = (\rho - 1)$ y Δ , es el operador de la primera diferencia.

El test consiste en probar:

Ho: $\delta = 0$. Si $\delta = 0$, entonces $\rho = 1$; es decir, la variable Y_t tiene una raíz unitaria, lo que significa que la serie es no estacionaria

Hi: Y_t es estacionaria

La contribución de DF fue determinar la nueva distribución necesaria para las pruebas y crear las tablas, para eliminar la posibilidad de que $\delta > 0$ porque en ese caso $\rho > 1$, y si fuera así, entonces la serie de tiempo subyacente sería explosiva.

Nota: Test valido solo para variables AR(1).

b. Test Dickey-Fuller aumentado (DFA): al llevar a cabo la prueba DF, se supuso que el término de error μ_t no estaba correlacionado. Pero Dickey y Fuller desarrollaron una prueba cuando dicho término si está correlacionado, la cual consiste en aumentar los valores rezagados de la variable dependiente ΔY_t .

$$Y_t = \alpha + \beta t + \sum \rho_i Y_{t-1} + \mu_t$$

En la DFA se sigue probando $\delta = 0$ y además esta prueba sigue la misma distribución asintótica que el estadístico DF, por lo que se pueden utilizar los mismos valores críticos.

c. Test Phillips-Perron (PP): una importante suposición de la prueba DF es que los términos de error μ_t están distribuidos de manera idéntica e independiente. Phillips y Perron utilizan métodos estadísticos no paramétricos para evitar la correlación serial en los términos de error, sin que añadan términos de diferencia rezagados.

ANÁLISIS DE COINTEGRACIÓN

Supóngase que se consideran las series de tiempo “Gasto” e “Ingreso” de Venezuela. Si estas series se someten de manera individual a un análisis de raíz unitaria, se encontrará que ambas son $I(1)$; es decir, tienen una raíz unitaria. Por lo tanto, supóngase que se hace la regresión del “Gasto” sobre el “Ingreso” de la siguiente forma:

$$Gasto_t = \beta_0 + \beta_1 Ingreso + \mu_t$$

Esto se expresa como:

$$\mu_t = Gasto_t - \beta_0 - \beta_1 Ingreso$$

Supóngase que ahora se somete μ_t a un análisis de raíz unitaria y se descubre que es estacionaria; es decir, es $I(0)$. Bajo esta situación, las variables individualmente tienen tendencias estocásticas y su combinación lineal es $I(0)$; por lo tanto se podría decir que la combinación lineal cancela las tendencias estocásticas en las dos series. Como resultado, la regresión planteada en este caso es **cointegrada**.

En términos económicos, dos variables serán cointegradas si existe una relación a largo plazo, o de equilibrio, entre ambas.

La regresión estimada se conoce como **regresión cointegrante** y el parámetro estimado se conoce como **parámetro cointegrante**. El concepto de cointegración puede extenderse a un modelo de regresión que contenga k regresoras.

Metodología Engle-Granger

Este método se basa en aplicar la prueba de raíz unitaria DF o DFA, de la siguiente manera:

- 1ro. Determinar el orden de integración de las variables. En caso de que alguna sea de orden superior, diferenciar la serie.
- 2do. Estimar el modelo aplicando el método de Mínimos Cuadrados Ordinarios
- 3ro. Verificar que los residuos sean estacionario; es decir, $I(0)$.