EasyReg: Aplicaciones para un curso de econometría

Julio César Alonso y Paul Semaán

UNIVERSIDAD ICESI

EASYREG

APLICACIONES PARA UN CURSO DE ECONOMETRÍA

Julio César Alonso C.

Paul Semaán R.

EasyReg: Aplicaciones para un curso de econometría / Julio César Alonso, Pául Semaán R.
1 ed. Cali: Universidad Icesi 2010.
242 p; ISBN 978-958-8357-29-4
1. Econometría 2. EasyReg 3. Problemas econométricos

EasyReg: Aplicaciones para un curso de econometría
Colección "Discernir"
Universidad Icesi

http://www.icesi.edu.co/investigaciones_publicaciones/libros/easyreg_aplicaciones_econometria

Rector
Francisco Piedrahita Plata

Secretaria General
María Cristina Navia

Director Académico
José Hernando Bahamón

Directora de Investigaciones y Publicaciones
Diana Carolina Mora

Diseño de carátula
Gustavo Andrés Álvarez

Diseño de textos
Paul Semaán Riascos

Impreso en Cali-Colombia

A.A. 25608 Unicentro
Tel. 5552334 Ext. 8404
Fax: 5551706
E-mail: publicaciones-icesi@icesi.edu.co
Cali, Colombia

Primera edición, julio de 2010

ISBN 978-958-8357-29-4

Índice general

Prefacio

El análisis empírico de la realidad a la luz de la teoría es cada día más una parte fundamental en la formación de un economista. La microeconomía, la macroeconomía y todo el andamiaje teórico de los programas de pregrado son complementados con la econometría para constituir la "caja de herramientas" de un economista. Así la econometría no se puede entender como un fin, sino como un medio para encontrar las regularidades presentes en una realidad.

Este libro pretende cumplir dos funciones. La primera es convertirse en un documento de consulta para economistas que desean emplear métodos econométricos convencionales y el paquete gratuito EasyReg. La segunda función es complementar un libro de texto convencional de econometría brindando el puente entre la teoría econométrica y la práctica econométrica mediante el uso de paquete econométrico gratuito. Así, este texto no pretende ser un sustituto de los manuales de econometría actualmente disponibles en el mercado, sino una complemento de estos.

EasyReg: Un paquete econométrico fácil de usar y gratuito

En la práctica, emplear métodos econométricos no sólo implica conocer las herramientas teóricas sino también escoger un Software para realizar los respectivos cálculos. En la actualidad, en el mercado están disponibles un gran número de Software licenciado que presentan diferentes opciones al usuario. Entre los paquetes licenciados más empleado se encuentran EViews, JPM, LIMDEP, Matlab, Minitab, RATS, SAS, Shazam, SPSS, Stata y TSP.

Por otro lado, en la Web están disponibles algunos paquetes econométricos gratuitos, como EasyReg y R, que permiten el uso de métodos econométricos para cualquier usuario que cuente con un PC. Las licencias cada vez son menos una barrera que impide el uso de métodos econométricos. El software gratuito permite a los docentes brindarle a los estudiantes la facilidad de instalar en sus propios computadores un software econométrico, sin violar ningún derecho de autor. Las prácticas, tareas o proyectos pueden ser desarrollados por los estudiantes en cualquier PC sin violar leyes.

Entre los paquetes gratuitos sobresale EasyReg International desarrollado por el profesor Herman Bierens de la Universidad de Pennsilvania State. Este paquete es amigable para el usuario al ser una aplicación que se maneja únicamente por menús, señalando y haciendo clic por medio de un ratón. Así, no es necesario memorizar comandos, ni aprender programación para obtener cálculos econométricos relativamente sofisticados.

Su carácter gratuito sin lugar a dudas lo convierte en una buena opción de paquete econométrico para la instrucción de los métodos econométricos fundamentales, así como su uso en la investigación, en países en desarrollo. Pero el costo del paquete no es el único atractivo. Una ventaja de este paquete, frente a paquetes similares como el Eviews 7 (disponible a un costo de 425 dólares), es que EasyReg está escrito tanto para principiantes como para investigadores que deseen emplear métodos estadísticos relativamente avanzados. De hecho, existe una opción por medio de la cual el usuario puede habilitar el nivel de complejidad deseada: estudiante principiante, estudiante intermedio, estudiante avanzado o investigador. El paquete está equipado con rutinas que permiten graficar la información y calcular estadísticas descriptivas como la media, la varianza, percentiles, etc. En este aspecto, el paquete no tiene nada que envidiarle a paquetes similares como Eview, SPSS y JPM. Tal vez la única desventaja que presenta este paquete, frente a los mencionados anteriormente, es que los gráficos no pueden ser manipulados y su aspecto no es el mejor si se desea una gráfica para ser publicada en un artículo.

Por otro lado, EasyReg proporciona una serie de métodos univariados, uniecuacionales y multiecucionales pre-programados más allá de los convencionales para un paquete de este tipo. Las opciones disponibles para efectuar análisis univariado de series de tiempo incluyen estimación de modelos ARMA, GARCH, estimación de periodogramas, pruebas de raíces unitarias como las pruebas de Dickey-Fuller aumentado, Phillips-Perron, Bierens, Bierens-Guo y KPSS.

Los métodos de análisis uniecuacional disponibles en EasyReg no sólo se limitan a lo estándar como la estimación de modelos de regresión lineal múltiple por el Método de Mínimos Cuadrados Ordinarios, sino que también incluye opciones para estimar modelos con variables dependientes discretas como modelos Logit, Probit, regresión de Poison, modelos de regresión con datos censurados (Modelo Tobit), Regresión por cuantiles y Mínimos Cuadrados bi-etápicos.

En cuanto a los métodos para estimación de sistemas de ecuaciones, EasyReg permite estimar modelos por medio del método de Momentos Generalizados incluyendo el método SUR (Regresiones aparentemente no relacionadas), estimar modelos VAR (tanto estructurales como sin restricciones) y sus respectivas funciones de impulso-respuesta, efectuar análisis de cointegración por medio de la prueba de Johansen y estimación de modelos de corrección de errores. Adicionalmente, es posible efectuar la prueba de cointegración no paramétrica de Breitung y Bierens. Finalmente, EasyReg permite estimar modelos no lineales definidos por el usuario empleando los métodos de mínimos cuadrados, máxima verosimilitud y generalizado de Momentos. Esta opción no está disponible en Eviews.

Una cualidad de EasyReg, presente en todos los métodos anteriormente mencionados, es la sencillez y facilidad de lectura de los resultados arrojados por el paquete. En especial, EasyReg incluye regularmente la hipótesis nula de la mayoría de las pruebas, los valores críticos relevantes y la decisión a tomar. Esta característica convierte al paquete en una excelente herramienta de enseñanza de la econometría, tanto a nivel introductorio como intermedio. La simplicidad de operación de este programa está presente en todo el proceso de estimación; hecho que permite quitar la presión al estudiante o investigador de manejar lenguajes sofisticados y rutinas de programación para obtener los resultados deseados. Así, el nexo entre los textos, el "tablero" y la práctica se realiza de una manera natural para el estudiante, facilitando el proceso de aprendizaje. Y, ¡lo mejor de todo para el estudiante es el precio del paquete!

Los economistas enseñamos que no hay almuerzo gratis, pero este software claramente contradice este principio. EasyReg es una excelente solución tanto para la docencia como para la investigación. Sin lugar a dudas, este paquete será cada día más popular para la instrucción de la econometría de pregrado. No solamente por sus virtudes anteriormente descritas, sino también porque es posible brindarle a los estudiantes una copia legal del paquete sin ningún costo, e instalar en todas las salas de cómputo de las instituciones las copias necesarias.

Definitivamente EasyReg Internacional es un paquete digno de ser considerado antes de invertir su dinero en paquetes similares como Eviews, Microfit, JPM4, etc. Esta aplicación está disponible para ser descargado directamente de la página oficial http://econ.la.psu.edu/ hbierens/EASYREG.HTM y de diferentes sitios alrededor del mundo, incluyendo uno en Colombia (http://www.icesi.edu.co/ jcalonso).

Organización de esta obra

La obra está compuesta de tres partes. En la primera se discute el modelo de regresión múltiple y la inferencia asociada a este. La segunda parte discute los problemas econométricos más comunes presentes en el modelo de regresión múltiple. En cada capítulo de esta sección se discuten problemas como la multicolinealidad, heteroscedasticidad y autocorrelación, las implicaciones de dichos problemas, las pruebas estadísticas que permiten detectarlos dichos problemas y cómo solucionarlos. Finalmente, la última sección discute la estimación de modelos de ecuaciones simultáneas y modelos con una variable dicotómica como variable dependiente.

Cada capítulo contiene tres partes. La primera discute rápidamente el tema a un nivel teórico empleando una aproximación matricial; esta revisión se puede entender como un breve resumen que complementa la lectura de un libro de texto introductorio en el tema. En ningún momento pretende reemplazar el libro de texto seleccionado. La segunda sección presenta paso a paso el procedimiento para aplicar en EasyReg los conceptos revisados de manera teórica en la sección anterior. Para la explicación paso a paso de cómo implementar los modelos econométricos se emplean datos reales frutos de investigaciones desarrolladas por profesores del departamento de Economía de la Universidad Icesi. Estos datos se encuentran en la página Web del libro, para permitir al lector seguir paso a paso la exposición de la obra. La última parte de cada capítulo presenta ejercicios que permitirán al lector practicar los temas discutidos. Las respuestas a los ejercicios planteados se presentan al final del libro.

Esta obra en la docencia de la econometría en pregrado

Es innegable la gran importancia que tiene la econometría en el ejercicio profesional de los economistas que se dedican a la investigación empírica. En estos momentos es difícil pensar proyectos de investigación aplicada que no empleen alguna de las técnicas estadísticas comunes en la profesión. No obstante, es importante anotar que los economistas que se dedican a la investigación son la minoría, y con un nivel de formación igual o superior a una maestría. Si bien no existen cifras claras, intuitivamente parece evidente que buena parte de los egresados de los programas de economía del país se dedican a actividades diferentes a la investigación.

El aporte de la enseñanza de la econometría en un currículum de pregrado radica en dos puntos de vital importancia: i) introducir una herramienta metodológica, con sus bondades y limitaciones, que permite cuantificar los modelos teóricos estudiados en los otros cursos y ii) introducir al futuro economista al lenguaje empleado en los estudios de corte econométrico.

Es imposible pretender brindar en uno o dos cursos de pregrado las herramientas necesarias para que el estudiante se convierta en un economista con capacidad de enfrentar los problemas relativamente comunes que se presentan en la estimación de modelos econométricos. Así los cursos de econometría de pregrado no pueden tener como objetivo formar expertos econometristas.

Dos objetivos más acordes con la formación de pregrado pueden ser i) brindar al estudiante la capacidad de interpretar los resultados de un modelo estimado por un tercer y ii) brindar la capacidad de realizar las preguntas relevantes para determinar la pertinencia estadística y económica de los modelos

estimados. En este orden de ideas, la obra está más dedicada a facilitar el proceso de estimación de los modelos. Con esto pretendemos "liberar" el espacio y el tiempo necesarios para que en los cursos sea posible discutir el cómo se interpretan los resultados y no se dedique tanto esfuerzo en la discusión del cómo se calcula.

Este libro pretende ser una compañía a los textos tradicionales de econometría al presentar rápidamente los conceptos teóricos y mostrar paso a paso cómo deberían ser estimados los modelos econométricos.

Agradecimientos

Tras ocho años de emplear parte del material que conforma esta obra como notas de clase de un curso introductorio de Econometría en la Universidad Icesi de Cali, decidimos convertir estas notas en una obra autocontenida. Los contenidos de los capítulos y su arquitectura, son producto de los comentarios valiosos de cerca de 300 estudiantes de este curso. En especial los monitores del curso de econometría han brindado comentarios muy útiles que nos permiten estar cómodos con esta versión del documento. Nos sería imposible iniciar esta obra sin destacar monitores sobresalientes como Sumie Tamura, Sebastián Solanilla, Margareth González, Ana María Lotero, Carlos Patiño, Francisco Quevedo, Stephany Vergara, Mauricio Arcos, David Valencia, Ana Isabel Gallego y Manuel Serna. Nuestros agradecimientos eternos a su colaboración y comentarios.

Parte I

El modelo de regresión múltiple

CAPÍTULO 1

Modelo de regresión múltiple

Objetivos del capítulo

El lector, al finalizar este capítulo, estará en capacidad de:

- Estimar un modelo lineal con más de una variable aleatoria empleando EasyReg .
- Identificar los coeficientes estimados por EasyReg.
- Transformar y crear variables en EasyReg.
- Interpretar los componentes de las tablas de salida que proporciona EasyReg.

1.1. Introducción

La teoría económica nos provee con diferentes relaciones entre variables económicas; por ejemplo, sabemos que la demanda de un bien Q depende del precio de los bienes sustitutos que pueda tener éste, el precio de los bienes complementarios, del nivel de ingresos, de los gustos y del número de consumidores. Estas relaciones teóricas se pueden representar funcionalmente como:[1]

$$Q = Q_x\left(p_x, p_{comp}, p_{sust}, I\right) \tag{1.1}$$

Además, la teoría nos permite conocer cuál sería el efecto de cambios de cada una de estas variables independientes, ceteris paribus, sobre las cantidades demandadas. En otras palabras, la teoría económica permite conocer el signo de $\frac{\partial Q_x}{\partial p_x}$, $\frac{\partial Q_x}{\partial p_{comp}}$, $\frac{\partial Q_x}{\partial p_{sust}}$, etc. Pero en general, la teoría no nos da indicios de la forma de la función de demanda; es decir, la teoría no puede brindar elementos para decidir entre diferentes formas funcionales. Por ejemplo,

$$\begin{aligned} \mathrm{Q}_x\left(p_x, p_{comp}, p_{sust}, I\right) &= Ap_x^{\alpha} + \frac{1}{p_{comp}^{\beta} + C} + \ln\left(\varphi p_{sust}\right) + \alpha I \\ \mathrm{Q}_x\left(p_x, p_{comp}, p_{sust}, I\right) &= \beta_0 + \beta_1 p_x + \beta_2 p_{comp} + \beta_3 p_{sust} + \beta_4 I \end{aligned} \tag{1.2}$$

Así, la teoría económica provee las relaciones funcionales teóricas, pero en general no provee la forma funcional explícita. Aún más, el carácter de la mayoría de las relaciones funcionales son determinísticas.[2] Por tanto relaciones funcionales como la expresada en la ecuación 1.1 corresponden a modelos matemáticos (exactos) y no estadísticos.

En la práctica se reconoce que las relaciones entre las variables independientes y la dependiente no tienen por qué ser exactas y se incluye un término aleatorio de error en los modelos económicos a estimar.[3] La inclusión de este término de error se justifica de diferentes formas. Por ejemplo,[4]

- Las respuestas humanas individuales a diferentes incentivos no son exactas y por tanto no son predecibles con total certidumbre; aunque se espera que en promedio si lo sean.
- En general, es imposible pretender que un modelo recoja todas y cada una de las variables que afectan directamente una variable, pues precisamente un modelo económico es una simplificación de la realidad y por tanto omite detalles de ella. Es importante anotar que en algunas ocasiones las variables que se omiten son conocidas, pero el investigador no cuenta con información para medir esas variables, por tanto deben ser omitidas.
- La variable dependiente puede estar medida con error, pues en la práctica los agregados económicos normalmente son estimados a partir de muestras. El error de medición en la variable dependiente estará recogido en el término de error, pues nuestro modelo no pretenderá explicar el error de medición sino el comportamiento promedio del agregado económico.

[1] Recuerde que antes de la lectura de este capítulo es importante estudiar en su libro texto los capítulos relacionados con el modelo lineal y los estimadores de Mínimos Cuadrados Ordinarios.

[2] No incluyen una variable aleatoria.

[3] Noten que si no existiera un término aleatorio en nuestro modelo, entonces estaríamos hablando de modelos determinísticos y no requeriríamos de métodos estadísticos para determinar los parámetros del modelo.

[4] Para ver más justificaciones para la inclusión del término de error el lector puede consultar Gujarati, Damodar. 1997. Econometría: McGraw Hill. Pág. 38.

- En algunas oportunidades las relaciones entre variables anunciadas por la teoría económica son producto de un esfuerzo de resumir un conjunto de decisiones individuales. Así como las decisiones individuales son diferentes de individuo a individuo, cualquier intento por estimar estas relaciones a nivel agregado será simplemente una aproximación; por tanto la diferencia entre esta aproximación y el valor real será atribuida al término de error.

Entonces, todo modelo econométrico poseerá una parte aleatoria y una que no lo es (parte determinísticas). Por ejemplo, si se considera la relación funcional expresada en la ecuación 1.1, el correspondiente modelo estadístico corresponderá a $Q = Q_x\left(p_x, p_{comp}, p_{sust}, I\right) + \varepsilon$, donde $Q_x\left(p_x, p_{comp}, p_{sust}, I\right)$ corresponde a la porción determinística del modelo y ε representa el término aleatorio de error.

Como se mencionó antes, generalmente la teoría económica brinda las relaciones funcionales y de causalidad entre diferentes variables económicas. Pero la teoría no provee explícitamente el tipo de relación funcional entre las variables explicativas y la variable dependiente. En la práctica, los investigadores adoptan como estrategia tomar una primera aproximación a las relaciones funcionales expresadas en la teoría por medio de relaciones lineales o linealizables. En este capítulo concentraremos nuestra atención en los modelos lineales. Es decir, modelos de la siguiente forma:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \tag{1.3}$$

En este modelo podemos distinguir varios componentes: la variable dependiente (Y_i), las variables independientes (X_{1i} y X_{2i}), los coeficientes o parámetros[5](β_0, β_1 y β_2), y el término de error (ε_i).

Antes de entrar en detalle, es importante aclarar qué se entiende por modelo lineal en este contexto. Para ser más precisos, cuando nos referimos a un modelo de regresión lineal, estamos hablando de un modelo que es lineal en sus parámetros y el término de error es aditivo. En otras palabras, los parámetros[6] están multiplicando a las variables explicativas o representan un intercepto. Formalmente, *un modelo se considera un modelo lineal si la variable dependiente se puede expresar como una combinación lineal de las variables explicativas y un término de error, donde los parámetros son los coeficientes de la combinación lineal.*

Por ejemplo, el modelo $Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$ es un modelo (estadístico) lineal, pues es lineal en los parámetros $\beta^T = (\beta_0, \beta_1, \beta_2)$; en este caso, estos representan un intercepto y pendientes (ver Ejemplo 1), y el término de error es aditivo. Por otro lado, un modelo como $Y_i = \beta_0 + (X_{1i})^{\beta_1} + \beta_2 X_{2i} + \varepsilon_i$ no es un modelo lineal, pues el modelo no es lineal respecto a β_1, el cual representa una potencia.

Ahora bien, es importante resaltar que un modelo econométrico puede ser lineal en los parámetros y tener un término aleatorio aditivo, pero no representar una línea recta, un plano o sus equivalentes en dimensiones mayores. En estos casos, aunque no se trata de un modelo lineal desde el punto de vista matemático, aún tenemos un modelo estadístico lineal (Ejemplo 1.2).

[5]Los parámetros corresponden a números (constantes) que describen la relación entre las variables independientes y la variable dependiente. Más adelante se ampliará esta idea.

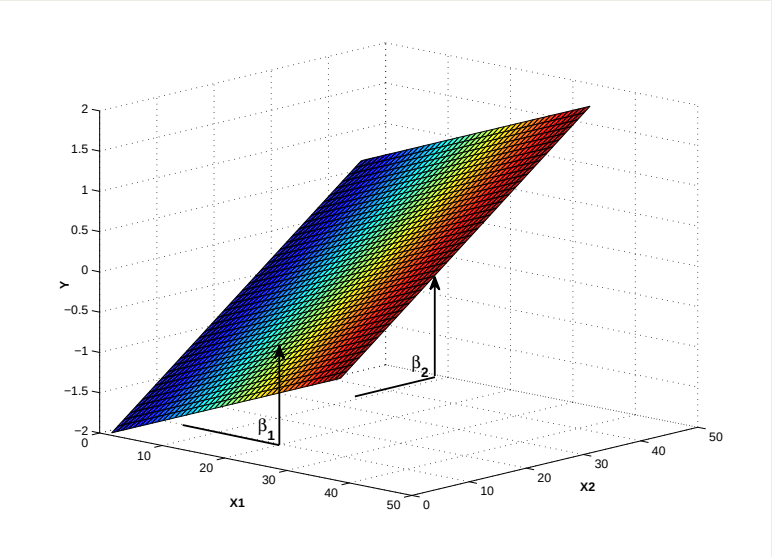
[6]A excepción de la varianza del término de error.

Ejemplo 1.1 Modelo de regresión lineal

Suponga la siguiente relación entre una variable dependiente (Y_i) y dos variables explicativas(X_{1i}, X_{2i}):

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i \tag{1.4}$$

donde ε_i es un término aleatorio de error. Entendamos primero la naturaleza de esta relación funcional omitiendo el término aleatorio de error. En el siguiente gráfico se puede observar la función $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$. Noten que $\frac{\partial Y}{\partial X_1} = \beta_1$, es decir el coeficiente asociado a la variable independiente X_1 corresponden al cambio en la variable dependiente cuando la otra variable independiente se mantiene constante, es decir la pendiente de la función con respecto al plano formado por los ejes de X_1 y Y. Similar interpretación posee $\beta_2 = \frac{\partial Y_i}{\partial X_{2i}}$.

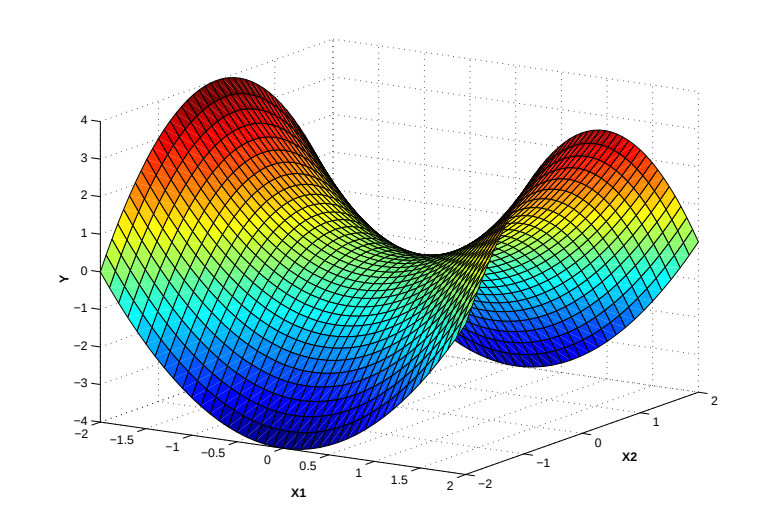


Ejemplo 1.2 Modelo intrínsecamente lineal

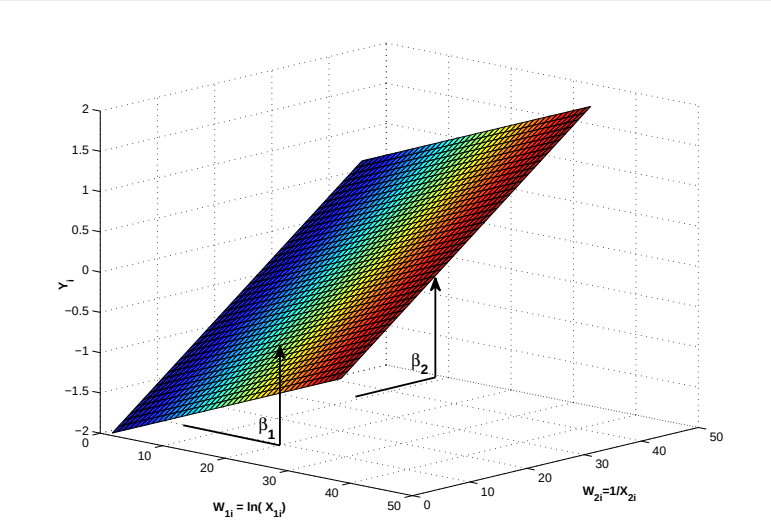
Suponga la siguiente relación entre una variable dependiente (Y_i) y dos variables explicativas(X_{1i}, X_{2i}):

$$Y_i = \alpha_1 + \alpha_2 \ln(X_{1i}) + \alpha_3 \left(\frac{1}{X_{2i}}\right) + \varepsilon_i \tag{1.5}$$

donde ε_i es un término aleatorio de error. En este caso, tenemos que el modelo es lineal en los parámetros α_1, α_2 y α_3, además el error es aditivo; por tanto este modelo es lineal. Es importante anotar que en este caso α_2 no es una pendiente, pues $\frac{\partial Y}{\partial X_1} = \frac{\alpha_2}{X_1}$; para α_3 ocurre algo similar. Por tanto este modelo no representa un plano (no es lineal desde el punto de vista matemático), como se puede observar en el siguiente gráfico en el cual se ha omitido el término de error, y por tanto solo se ha graficado la porción determinística del modelo



Pero esta función se puede expresar de tal forma que represente un plano; reparametrizando, podemos definir $W_{1i} = \ln(X_{1i})$ y $W_{2i} = \left(\frac{1}{X_{2i}}\right)$. Ahora, remplazando estas dos nuevas variables en el modelo original, tenemos que $Y_i = \alpha_1 + \alpha_2 W_{1i} + \alpha_3 W_{2i} + \varepsilon_i$. Si graficamos esta nueva función, ignorando el término de error, en el espacio (Y, W_1, W_2), se obtendrá lo siguiente:



Este modelo reparametrizado es un modelo lineal y representa un plano; y se puede estimar por medio de los métodos que estudiaremos en este capítulo.

Es importante anotar que la interpretación de los parámetros α_1 y α_2 es un poco diferente en este caso. Por ejemplo, α_2 representa el número de unidades en que cambiará la variable dependiente cuando W_1 cambia en una unidad, es decir cuando el $\ln(X_1)$ cambia en una unidad y no cuando X_1 cambia. Para interpretar α_2 en términos de X_1 es necesario derivar (1.5) con respecto a X_1: $\frac{\partial Y_i}{\partial X_{1i}} = \frac{\alpha_2}{X_{1i}}$. Manipulando algebraicamente tenemos $\frac{\frac{\partial Y_i}{\partial X_{1i}}}{X_{1i}} = \alpha_2$. Multiplicando a ambos lados por $\frac{1}{100}$, obtenemos

$\frac{\partial Y_i}{\frac{\partial X_{1i}}{X_{1i}}\times 100} = \frac{\alpha_2}{100}$. Es decir $\frac{\partial Y_i}{\Delta\%X_{1i}} = \frac{\alpha_2}{100}$. Por tanto, la interpretación de α_2 en términos de X_1 es la siguiente: cuando X_1 aumenta en uno porciento, entonces Y_i aumentará en $\frac{\alpha_2}{100}$ unidades.

Existen otros modelos especiales que no son lineales en sus parámetros, pero mediante reparametrizaciones[7] se convierten en modelos lineales. Estos se conocen con el nombre de *modelos linealizables* (o modelos intrínsecamente lineales) y también pueden ser estimados por los mismos métodos lineales de estimación de un modelo lineal.

Ejemplo 1.3 Función de producción Cobb-Douglas

Una función de producción Cobb-Douglas se define como:

$$Y = AX_1^{\alpha_1}X_2^{\alpha_2}X_3^{\alpha_3} \tag{1.6}$$

donde Y corresponde a la producción, A denota el estado del conocimiento tecnológico actual, X_j representa la cantidad del insumo j empleado en el proceso productivo y α_j son los parámetros a estimar. Claramente, la expresión anterior no corresponde a un modelo estadístico, pues no presenta término aleatorio alguno. Como se discutió anteriormente, existen diferentes justificaciones para incluir un término estocástico de error en las relaciones que nos brinda la teoría económica. Así, un modelo econométrico a partir de la función de producción Cobb-Douglas sería:

$$Y_i = AX_{1i}^{\alpha_1}X_{2i}^{\alpha_2}X_{3i}^{\alpha_3}\varepsilon_i \tag{1.7}$$

donde $i = 1, 2, ..., n$ y ε_i es un término aleatorio de error. Pero este modelo no es lineal, al no ser lineal en los parámetros. No obstante, éste se puede linealizar fácilmente; aplicando logaritmos a ambos lados de la expresión tendremos:

$$\ln(Y_i) = \ln(A) + \alpha_1\ln(X_{1i}) + \alpha_2\ln(X_{2i}) + \alpha_3\ln(X_{3i}) + \ln(\varepsilon_i) \tag{1.8}$$

Definimos: $\beta = \ln(A)$ y $\mu_i = \ln(\varepsilon_i)$. Entonces la expresión anterior se puede reescribir de la siguiente forma:

$$\ln(Y_i) = \beta + \alpha_1\ln(X_{1i}) + \alpha_2\ln(X_{2i}) + \alpha_3\ln(X_{3i}) + \mu_i \tag{1.9}$$

Un aspecto importante de este modelo es la interpretación singular de los coeficientes. En este caso, cada α_i corresponde a la elasticidad del producto con respecto al insumo i . Por otro lado, el intercepto β corresponde al valor esperado del logaritmo del valor de la producción cuando todos los insumos son iguales a una unidad. En otras palabras $A = e^{\beta}$. Así, el intercepto en este modelo tiene una interpretación difícil y en la mayoría de las aplicaciones carece de interpretación económica.

[7] Una reparametrización es un cambio de nombre de variables y/o parámetros del problema que mantiene la naturaleza de la relación inalterada

1.2. El modelo de regresión múltiple

Una vez se cuenta con un modelo lineal que representa la relación entre diferentes variables explicativas y la variable dependiente, se deseará conocer los valores de los parámetros del modelo. Para lograr este fin, se recopilan observaciones de las variables explicativas y de la independiente. En general, un modelo lineal múltiple para el cual se cuenta con n observaciones y $k-1$ variables explicativas está dado por:

$$\begin{array}{ccccccccccc} y_1 & = & \beta_1 & + & \beta_2 X_{21} & + & \beta_3 X_{31} & + \ldots + & \beta_k X_{k1} & + & \varepsilon_1 \\ y_2 & = & \beta_1 & + & \beta_2 X_{22} & + & \beta_3 X_{32} & + \ldots + & \beta_k X_{k2} & + & \varepsilon_2 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ y_n & = & \beta_1 & + & \beta_2 X_{2n} & + & \beta_3 X_{3n} & + \ldots + & \beta_k X_{kn} & + & \varepsilon_n \end{array} \tag{1.10}$$

Para simplificar y ahorrar espacio, escribiremos el modelo 1.10 de una forma más abreviada, de tal modo que sólo describamos la observación *i-ésima* del modelo. Es decir:

$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \varepsilon_i \tag{1.11}$$

para $i = 1, 2, ..., n$. Otra forma de expresar el mismo modelo 1.11 de manera aún más abreviada es empleando matrices. Sean:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}_{k\times 1} \quad X = \begin{bmatrix} 1 & X_{21} & X_{31} & \cdots & X_{k1} \\ 1 & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix}_{n\times k} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

Entonces el modelo 1.11 se puede expresar de forma matricial así:

$$\mathbf{y_{n\times 1}} = \mathbf{X}_{n\times k}\beta_{k\times 1} + \varepsilon_{n\times 1} \tag{1.12}$$

Recuadro 1.1 Terminología de la regresión múltiple

$\boldsymbol{y}$	$\boldsymbol{X_2, X_3, ..., X_k}$
Variable dependiente	Variables independientes
Variable explicada	Variables explicativas
Variables de respuesta	Variables de control
Variables predicha	Variables predictoras
Regresando	Regresores

1.2.1. Supuestos

Es importante estudiar en detalle el vector de errores ε. En general, esperaremos que el error no sea predecible y por tanto trataremos de evitar cualquier componente determinístico o comportamiento sistemático del error. Así, asumiremos que en promedio el término de error es cero. En otras palabras, el

valor esperado de cada término de error es cero. En caso que el modelo incluya un intercepto, cualquier componente determinístico del error es capturado por el intercepto. Formalmente, asumiremos que: $E[\varepsilon] = 0$ para todo $i = 1, 2, ..., n$, o en forma matricial, $E[\varepsilon_i] = \mathbf{0}$.

Otro supuesto importante para garantizar que los errores son totalmente impredecibles es que cada uno de los errores sea linealmente independientes de los otros. En caso de existir dependencia lineal, habrá una forma de predecir errores futuros a partir de la historia de los errores. Por tanto, se asumirá que $E[\varepsilon_i \varepsilon_j] = E[\varepsilon_i] E[\varepsilon_j] = 0$, esto equivale a:

$$Cov(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i \varepsilon_j] - E[\varepsilon_i] E[\varepsilon_j] = 0$$

A este supuesto se le conoce como el supuesto de no autocorrelación entre los errores.

Finalmente, asumiremos que cada uno de los errores tiene la misma varianza, es decir se asumirá que:

$$Var[\varepsilon_i] = \sigma^2$$

para todo $i = 1, 2, ..., n$. Este supuesto se conoce como homoscedasticidad.

En resumen, asumiremos que el error cumple con los siguientes supuestos: 1) media cero, 2) varianza constante, y 3) independencia lineal entre los errores. Estos supuestos se acostumbran resumir de diferentes formas, por ejemplo se dirá que los errores están linealmente independientemente distribuidos con media cero y varianza constante, denotado por $\varepsilon_i \sim \text{l.i.d.}\left(0, \sigma^2\right)$. Otra forma de escribir estos supuestos de forma matricial es $\varepsilon \sim \left(0, \sigma^2 I_n\right)$, pues:

$$Var[\varepsilon] = \begin{bmatrix} Var[\varepsilon_1] & Cov(\varepsilon_1, \varepsilon_2) & \cdots & Cov(\varepsilon_1, \varepsilon_n) \\ Cov(\varepsilon_2, \varepsilon_1) & Var[\varepsilon_2] & \cdots & Cov(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(\varepsilon_n, \varepsilon_1) & Cov(\varepsilon_n, \varepsilon_2) & \cdots & Var[\varepsilon_n] \end{bmatrix}$$

$$Var[\varepsilon] = \begin{bmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{bmatrix} = \sigma^2 \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{pmatrix} = \sigma^2 \mathbf{I_n} \tag{1.13}$$

Por otro lado, asumiremos que la información aportada por cada una de las variables explicatorias al modelo es relevante. Es decir, no habrá ningún tipo de relación lineal entre las variables explicativas; pues en caso que una variable de control se pudiera expresar como una combinación lineal de otras variables explicativas, la información de la primera variable sería irrelevante pues ya está contenida en las otras. Así, asumiremos que $X_2, X_3, ..., X_k$ son linealmente independientes.

También, asumiremos que $X_2, X_3, ..., X_k$ son variables no estocásticas, pues se espera que el proceso de muestreo implícito en el modelo $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ debe poderse repetir numerosas veces y las variables exógenas no deben cambiar pues corresponden al diseño de nuestro "experimento". Así, se asumirá que $X_2, X_3, ..., X_k$ *son determinísticas*[8] *(no aleatorias) y linealmente independientes sí.*

Otros supuestos implícitos en nuestro modelo de regresión lineal $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ es que hay en efecto una *relación lineal* entre Y y $X_2, X_3, ..., X_k$. Asimismo, se supone que esta *relación es estable entre observaciones*; es decir, los parámetros del vector β son constantes a lo largo de la muestra.

[8]El supuesto de que las variables explicativas son determinísticas es un supuesto que se puede levantar sin muchas implicaciones. Pero por simplicidad, emplearemos este supuesto a lo largo del libro, a menos que se exprese lo contrario.

Finalmente, otro supuesto implícito en nuestro modelo de regresión lineal es la existencia de una causalidad unidireccional. Es decir nuestro modelo asume que las variables independientes $(X_2, X_3, ..., X_k)$ afectan a la variable dependiente (Y) y no al revés. Cabe mencionar que esta causalidad es provista por la teoría económica pues desde el punto de vista estadístico el modelo de regresión no tiene una connotación de causalidad asociada a él. Así, para el método estadístico es igualmente válido considerar una de las variables explicativas como variable dependiente y la variable como una variable explicativa. En la mayoría de las ocasiones, desde el punto de vista económico esta opción no es válida.

Recuadro 1.2 Supuestos del modelo de regresión múltiple

1. Relación lineal entre y y $X_2, X_3, ..., X_k$
2. $X_2, X_3, ..., X_k$ son fijas y linealmente independientes (la matriz X tiene rango completo)
3. el vector de errores ε satisface:

- Media cero $E[\varepsilon] = 0$
- Varianza constante
- No autocorrelación. Es decir: $\varepsilon_i \sim i.i.d\left(0, \sigma^2\right)$ o $\varepsilon_{n\times 1} \sim \left(\mathbf{0_{n\times 1}}, \sigma^2\mathbf{I_n}\right)$

1.2.2. Método de mínimos cuadrados ordinarios (MCO)

Como se mencionó anteriormente, dadas unas observaciones de la variable dependiente e independientes, normalmente se deseará conocer el valor de los coeficientes o parámetros (β). Para lograr acercarnos al valor poblacional desconocido de los coeficientes (β) se emplean estimadores (fórmulas) que responden a una idea plausible para "adivinar" el valor adecuado de éstos.

Una manera muy común de aproximarse a encontrar el "mejor" valor para los coeficientes poblacionales desconocidos (β) es minimizar la suma de los errores elevados al cuadrado;[9] este método se conoce con el nombre de Mínimos Cuadrados Ordinarios (MCO, o en inglés OLS). Intuitivamente, el método de MCO minimiza la suma de la distancia entre cada una de las observaciones de la variable dependiente $(\mathbf{y})$ y lo que el modelo "predice" $(\hat{\mathbf{y}} = \mathbf{X}\hat{\beta})$.

Formalmente, el estimador de MCO para el vector de coeficientes β, denotado por $\hat{\beta}$, se encuentra solucionando el siguiente problema:

$$\underset{\hat{\beta}}{Min}\left\{\left[\mathbf{y} - \mathbf{X}\hat{\beta}\right]^T\left[\mathbf{y} - \mathbf{X}\hat{\beta}\right]\right\} \tag{1.14}$$

donde $X\hat{\beta}$ corresponde al vector de valores estimados por el modelo para la variable dependiente, es

[9] Elevar al cuadrado el error tiene dos intenciones: 1) evitar que errores positivos y negativos se cancelen y 2) penalizar errores más grandes de manera más fuerte que errores pequeños.

decir el modelo estimado. Así, el estimador de MCO del vector de coeficientes β es:[10]

$$\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y} \tag{1.15}$$

La ecuación estimada estará dada por:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} \tag{1.16}$$

La diferencia entre el vector de valores observados de la variable dependiente $\mathbf{y}$ y los correspondientes valores estimados $\hat{\mathbf{y}}$ se denomina el *error estimado* o residuos y se denota por $\hat{\varepsilon}$; es decir:

$$\hat{\varepsilon} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta} \tag{1.17}$$

En el Apéndice 1.4 se discuten algunas propiedades importantes del vector $\hat{\varepsilon}$

Por otro lado, el estimador de MCO para la varianza σ^2 es:

$$s^2 = \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n-k} = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n-k} = \frac{\mathbf{y^T y} - \hat{\beta}^T\mathbf{X^T y}}{n-k} \tag{1.18}$$

Y el estimador de MCO para la matriz de varianzas y covarianzas de $\hat{\beta}$:

$$\widehat{Var\,[\beta]} = s^2\left(\mathbf{X^T X}\right)^{-1} \tag{1.19}$$

Puesto que $\left(\mathbf{X^T X}\right)^{-1}$ es simétrica, la matriz de varianzas y covarianzas también lo será. Esta matriz tiene la varianza de los estimadores en la diagonal principal y las covarianzas en las posiciones fuera de la diagonal principal. En otras palabras,

$$\widehat{Var\,[\beta]} = \begin{bmatrix} \widehat{Var\,[\beta_1]} & \widehat{Cov\,[\beta_1,\beta_2]} & \cdots & \widehat{Cov\,[\beta_1,\beta_k]} \\ & \widehat{Var\,[\beta_2]} & \cdots & \widehat{Cov\,[\beta_2,\beta_k]} \\ & & \ddots & \vdots \\ & & & \widehat{Var\,[\beta_k]} \end{bmatrix} \tag{1.20}$$

1.2.3. Propiedades de los estimadores MCO

El estimador de MCO de β es el estimador lineal insesgado con la mínima varianza posible, por esto, el estimador de MCO se conoce como el *Mejor Estimador Lineal Insesgado (MELI).*[11] Este resultado se conoce como el Teorema de Gauss-Markov (Recuadro 1.3).

Recuadro 1.3 Teorema de Gauss-Markov

Si se considera un modelo lineal $\mathbf{y}_{n\times 1} = \mathbf{X}_{n\times k}\beta_{k\times 1} + \varepsilon_{n\times 1}$ y se supone que:

- Las $X_2, X_3, ..., X_k$ son fijas y linealmente independientes (es decir X tiene rango columna completo y es una matriz no estocástica).

[10] En el Apéndice 1.2 se presenta la derivación de ésta fórmula.
[11] Una demostración de este resultado se presenta en el Apéndice 1.3

- El vector de errores ε tiene media cero, varianza constante y no autocorrelación. Es decir: $E[\varepsilon] = 0$ y $Var[\varepsilon] = \sigma^2 I_n$

Entonces, el estimador de MCO $\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$ es el *Mejor Estimador Lineal Insesgado (MELI)*

En la práctica, el Teorema de Gauss-Markov implica que si garantizamos que se cumplen los supuestos entorno al error y de una matriz de regresores con rango completo, entonces el estimador de MCO será MELI.

1.3. Ley de Okún en Colombia

En todos los capítulos de este libro, el lector encontrará ejercicios con datos de Colombia, los cuales serán resueltos paso a paso mediante el paquete estadístico EasyReg Internacional. En este capítulo explicaremos detenidamente desde cómo cargar los datos, hasta cómo estimar un modelo por MCO.

Para este ejercicio, nos basaremos en la ley de Okún para analizar la relación existente entre el PIB y la tasa de desempleo en Colombia. Empleamos datos de la tasa de desempleo y del PIB desde el primer trimestre de 2001 hasta el segundo semestre de 2006. Los datos están disponibles en el archivo adjunto (eregmult.xls).

Considere la siguiente relación dada por la teoría económica:

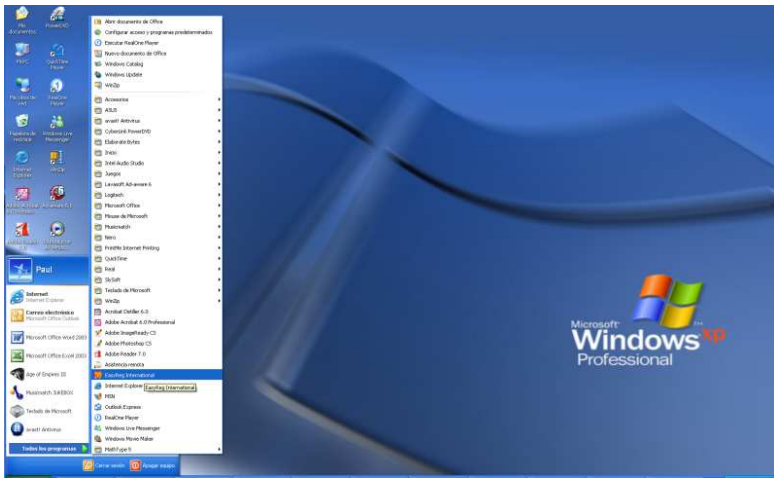
$$\Delta\,\%U_t = \beta_0 + \beta_1 \Delta\,\%PIB_t + \varepsilon_t \tag{1.21}$$

donde $\Delta\,\%U_t$ y $\Delta\,\%PIB_t$ corresponden al cambio porcentual en el periodo t de la tasa de desempleo y del PIB de Colombia, respectivamente. Dada la información contenida en el archivo adjunto, se desea estimar el modelo 1.21.

1.3.1. Lectura de datos

Las últimas versiones de EasyReg Internacional sólo leen datos que estén en el formato csv de Excel y archivos creados en el propio EasyReg. Si desea conocer cómo guardar archivos en formato csv desde Excel, por favor remítase al Apéndice 1.1.

Descargue el archivo con los datos a su disco duro. Guárdelos en cualquier directorio y abra el programa EasyReg Internacional haciendo clic en el icono de EasyReg en el menú de "Inicio/Programas". Durante el proceso de instalación del software un icono de atajo fue creado en su barra de "Inicio/programas".

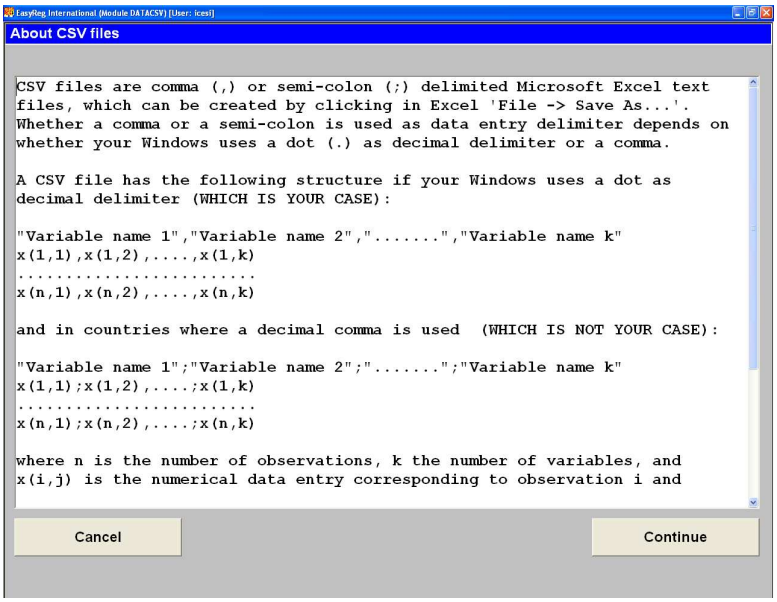


En caso que el icono de EasyReg Internacional no haya sido creado durante la instalación, busque en su disco duro la carpeta de archivos de programas y en dicha carpeta la subcarpeta EasyReg. En la subcarpeta EasyReg busque el archivo EasyReg.exe y haga doble clic en él.

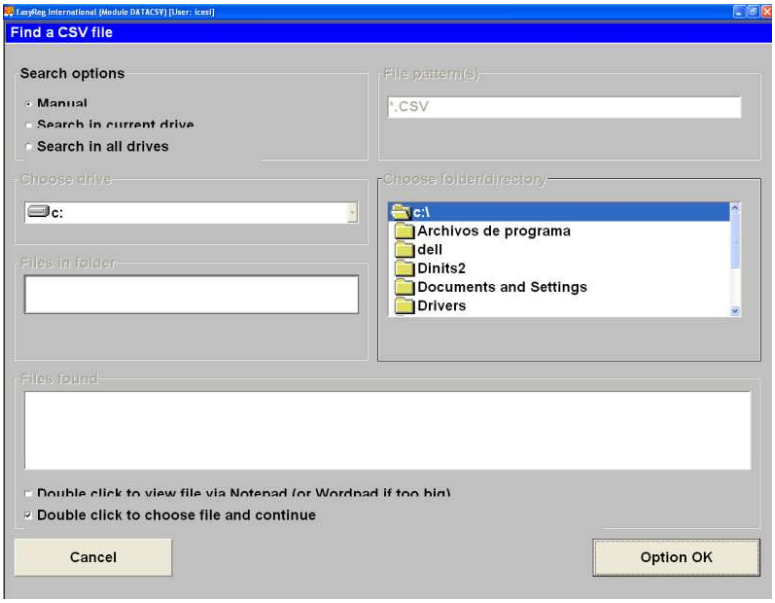
Una vez el programa EasyReg esté abierto, vaya al menú "File/Get data/Choose an Excel data file in CSV format".

Después de esto verá la siguiente ventana:[12]

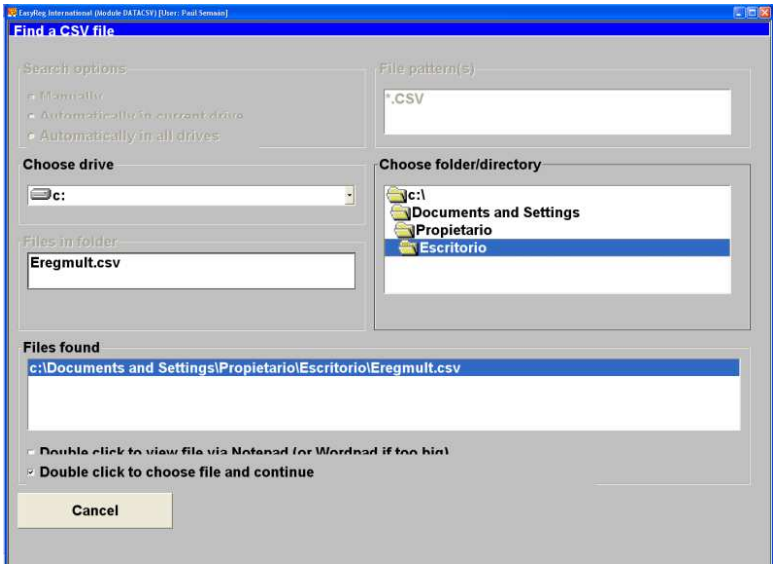
[12]Esta ventana es muy importante, pues brinda la información de cómo EasyReg está entendiendo el formato de los datos. Ver apéndice 1.1 para mayor detalle



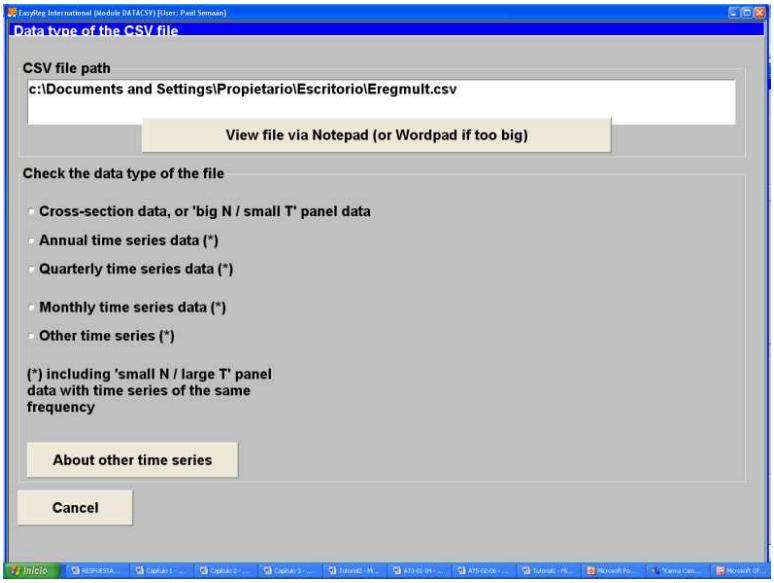
Haga clic en el botón de “Continue”. Esto lo llevará a la siguiente ventana:



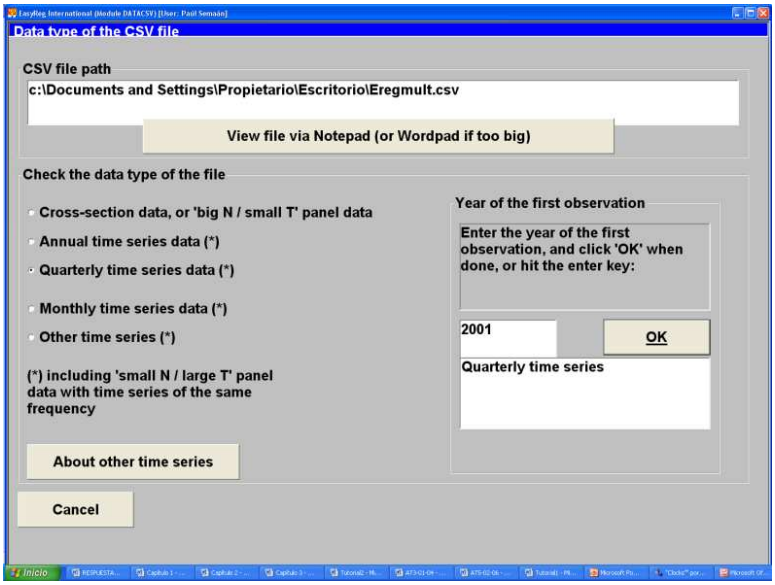
Haga clic en el botón “Option OK”. Ahora usted podrá escoger en la ventana “Choose drive” la unidad en la que grabó los datos; y en la ventana “Choose folder / directory” podrá identificar el lugar específico donde grabó los datos. Una vez identifique el lugar exacto donde se encuentran los datos verá en la ventana “Files in folder” el archivo que está buscando. Además, en la ventana “Files found” aparecerá el archivo deseado.

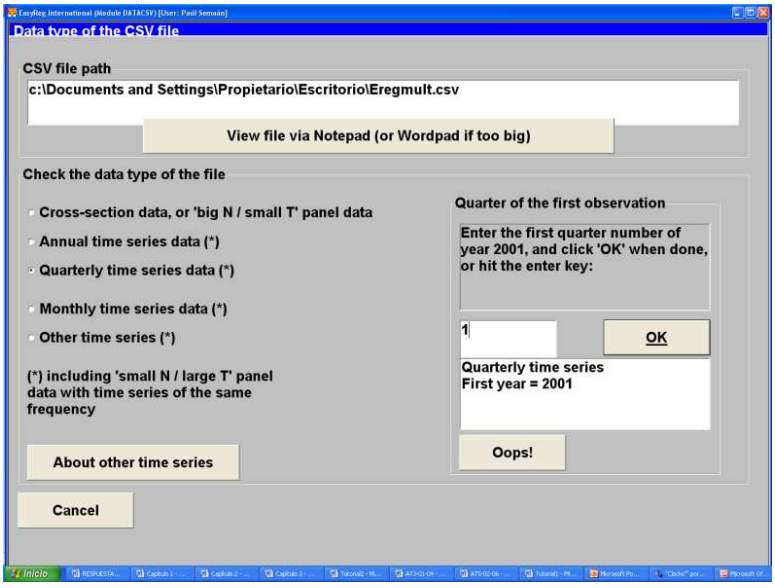


Haga doble clic en el archivo deseado en la ventana "Files Found". Verá la siguiente ventana:

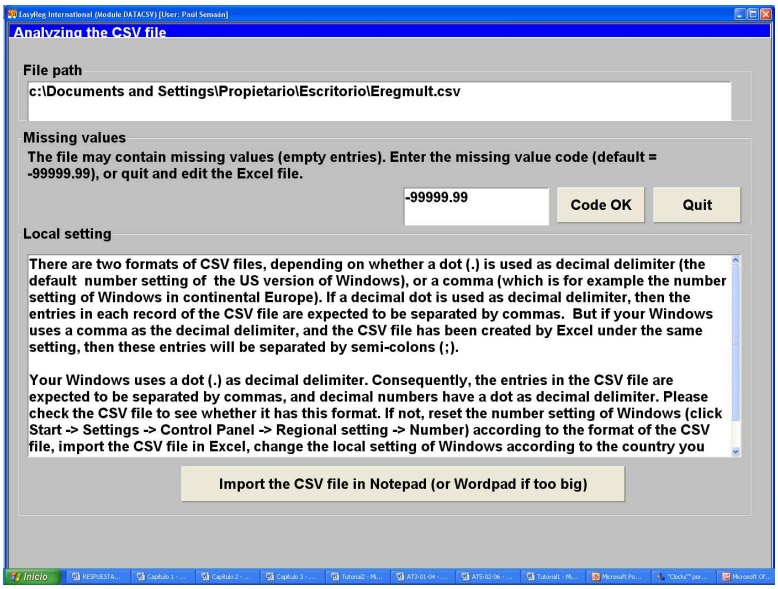


Ahora necesitamos decirle al programa la naturaleza de los datos que estamos empleando, es decir, datos anuales, trimestrales, de corte transversal, etc. En este caso, se trata de datos trimestrales desde 2001:1. Escoja la opción "Quarterly time series data (*)" haciendo clic sobre dicha opción. Inmediatamente, aparecerá una ventana preguntando por el año de la primera observación. Escriba en ella 2001, luego le preguntará el primer trimestre observado, escriba 1, haga clic en el botón "OK" y de nuevo haga clic en el botón "Confirm".

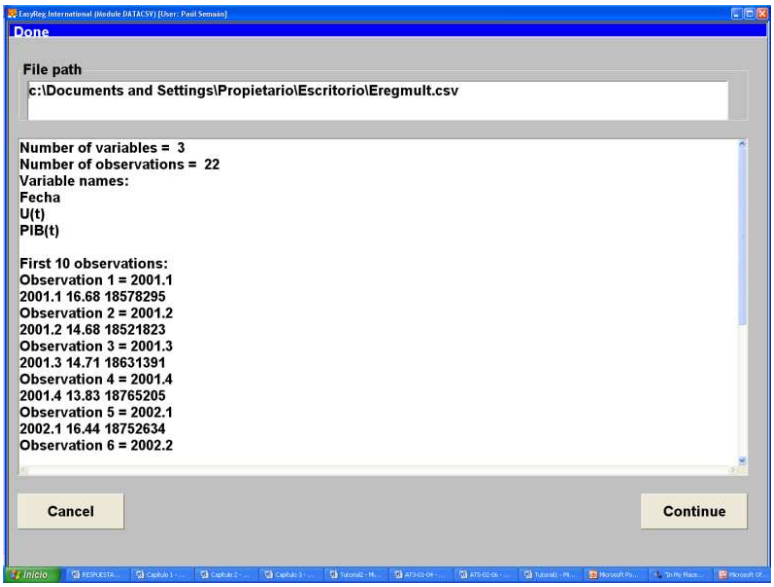




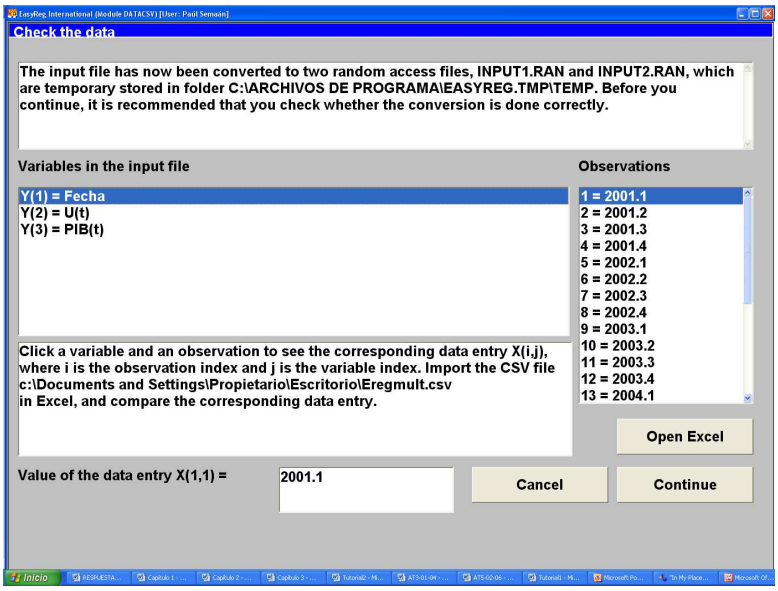
Después de esto, verá una nueva ventana que le pide confirmar la información anteriormente suministrada. Haga clic en el botón “Continue”. Ahora verá la siguiente ventana:



Esta ventana sirve para identificar la existencia de datos perdidos, que a veces en vez de dejar el espacio vacío se denotan por -99999,99. En este caso contamos con todos los datos, así no tenemos que preocuparnos por esto. Haga clic en "Code OK". Después de esto, usted deberá ver la siguiente ventana:

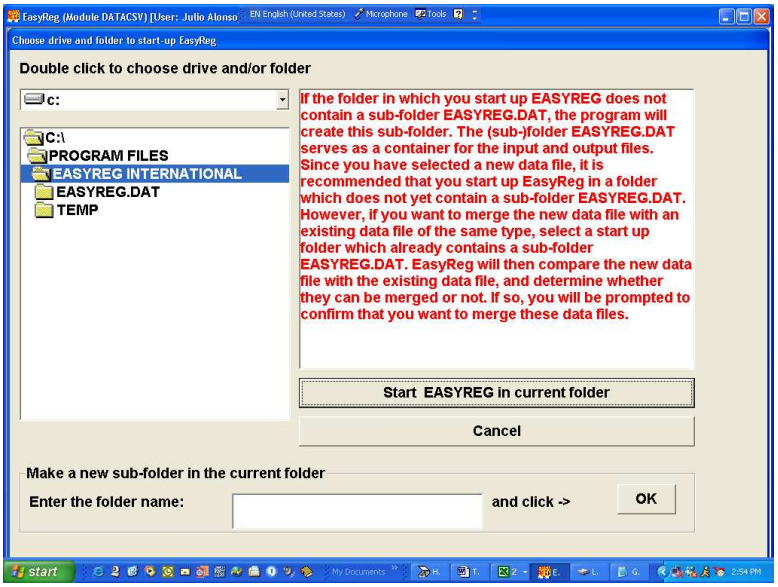


En caso que usted no observe esta ventana, es probable que EasyReg no haya leído correctamente los datos. Si éste es su caso, por favor repita los pasos anteriormente descritos y revise el Apéndice 1.1. Es probable que los datos no estén guardados en el formato deseado. En la ventana anterior se puede confirmar que los datos han sido leídos correctamente. Cerciórese que el número de variables ("Number of variables=") es 3 y que se han leído 22 observaciones ("Number of observations = 22"). Haga clic en el botón "Continue" y observará la siguiente ventana:



Esta ventana le permite revisar que los datos leídos sean los correctos. Los datos leídos son guardados en dos archivos temporales INPUT1.RAN e INPUT2.RAN, en el fólder C:\ARCHIVOS DE PROGRAMA\EASYREG INTERNATIONAL\TEMP. EasyReg creará una carpeta EasyReg.Dat que contiene los archivos con los datos INPUT1.RAN e INPUT2.RAN, y el archivo de Excel INPUT.CSV que también contiene los datos, además se crean los archivos OUTPUT.TXT, *.BMP (que contiene gráficos) y varios archivos *.TMP (archivos auxiliares) en el fólder donde está guardado el EasyReg. Estos archivos contendrán todos los resultados de sus cálculos.

Haga clic nuevamente en el botón “Continue” y en la ventana siguiente haga clic de nuevo en el botón “Continue”. Observará la siguiente ventana.



Haga clic en el botón “Start EasyReg in Current folder”. Esta opción le permite escoger el fólder en que EasyReg realizará los cálculos y almacenará todos los resultados y archivos temporales. Si usted está usando un computador en una sala de cómputo pública, lo mejor es que escoja la carpeta donde

está instalado EasyReg como el fólder de inicialización. La razón es que si escoge otro fólder, muy probablemente no tendrá permiso de escritura en otros fólderes, o en el peor de los casos, futuros usuarios del mismo PC no podrán correr el programa pues ellos no tienen acceso a los mismos fólderes a los que usted sí tiene acceso. Si está corriendo EasyReg en su computador personal, no hay ningún problema en escoger una carpeta diferente.

En la siguiente ventana haga clic en el botón "Overwrite or merge". En la ventana siguiente haga clic nuevamente el botón "Overwrite existing file". Ahora regresará a la ventana de inicio de EasyReg International. Los datos han sido leídos y están listos para empezar su análisis.

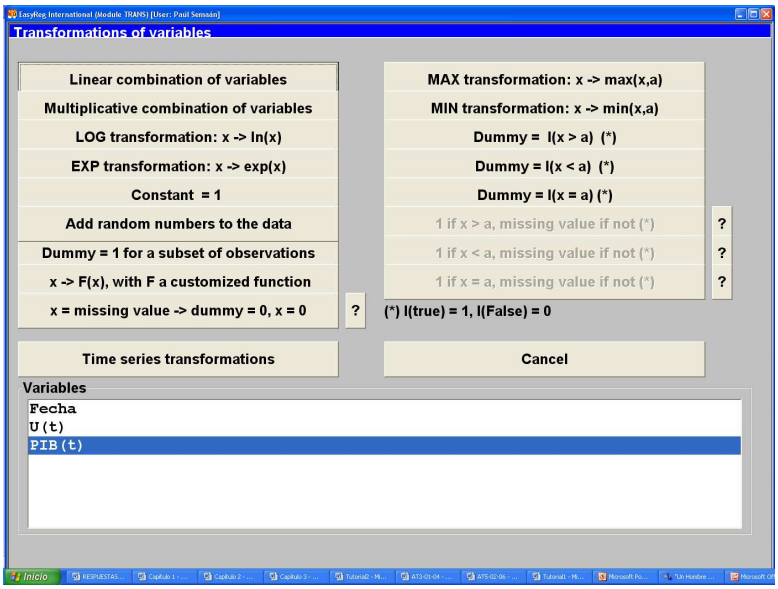
Si después de este paso observa un mensaje de error y el EasyReg se cierra sin darle ninguna opción, no se preocupe. Esto ocurre porque muy seguramente usted no tiene los permisos requeridos para acceder al fólder donde está instalado el EasyReg (Esto es muy común cuando usted trabaja en una sala pública de computadores). Si este es su caso, escoja un fólder diferente en el paso anterior. Un buen fólder para correr el EasyReg puede ser el fólder mis documentos, escritorio o en una memoria USB.

Finalmente, es importante notar que los últimos datos leídos por EasyReg quedan en memoria de tal forma que la próxima vez que usted abra el programa EasyReg, los últimos datos aún continúan en memoria y no necesitará cargarlos de nuevo.

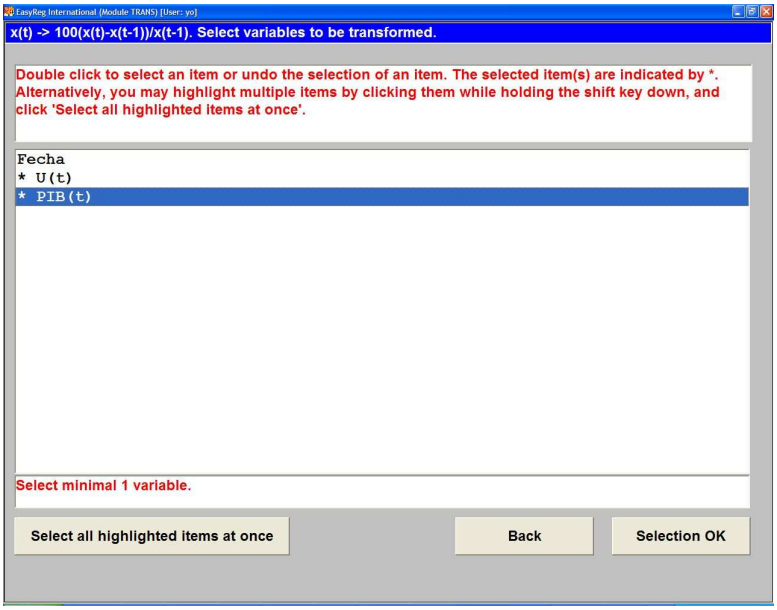
1.3.2. Transformación de datos

Los datos que acabamos de cargar en EasyReg no están expresados como un cambio porcentual, tal como estamos interesados en este ejercicio. Así pues, es necesario transformar la tasa de desempleo y el PIB en sus correspondientes variaciones porcentuales. Para efectuar este tipo de transformación, haga clic en "Menu / Input / Transform variables".

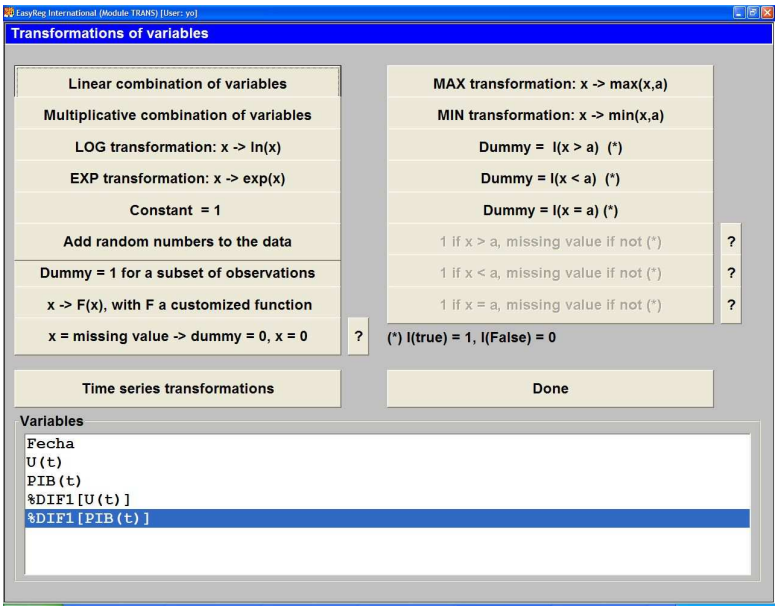
Aparecerá la siguiente ventana:



Para transformar las series en sus correpondientes crecimientos porcentuales, seleccione la opción "Time series transformation". Y luego "Percentage change: 100[x(t)-x(t-m)/x(t-m)]", dado que necesitamos los crecimientos de las series con respecto al periodo inmediatamente anterior; entonces conservamos el parámetro "Lag m = 1" que EasyReg usa por defecto. Pulse nuevamente "Percentage change: 100[x(t)-x(t-m)/x(t-m)]" y verá la siguiente ventana:



En esta ventana seleccione las variables PIB(t) y U(t) oprimiendo doble clic sobre cada una. Luego oprima "Selection O.K." y posteriormente "O.K". De esta manera usted ya habrá transformado las variables en crecimientos porcentuales.



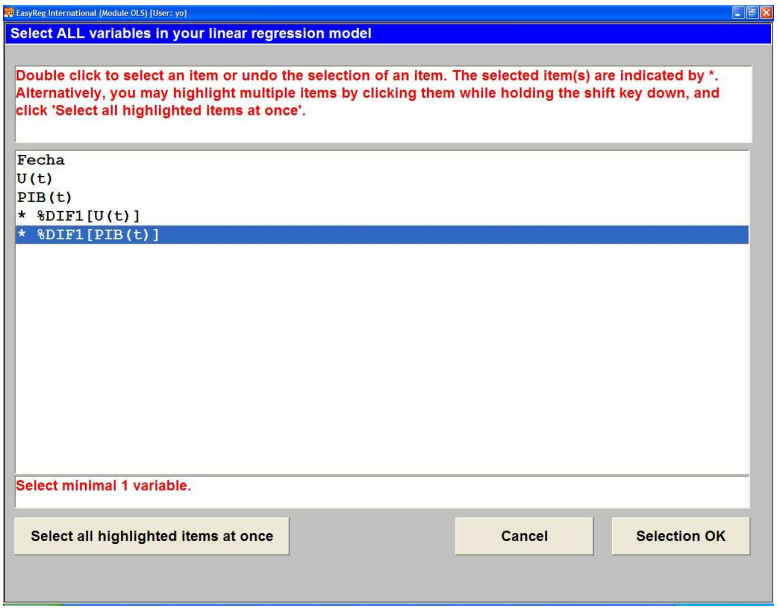
1.3.3. Estimación del modelo

Recordemos que se desea estimar el modelo:

$$\Delta\,\%U_t = \beta_0 + \beta_1 \Delta\,\%PIB_t + \varepsilon_t \tag{1.22}$$

Para estimar este modelo por el método de mínimos cuadrados ordinarios (una vez transformados los datos) debemos seguir los siguientes pasos. Primero, escoja la opción "Menu / Single equation models / Linear regression models".

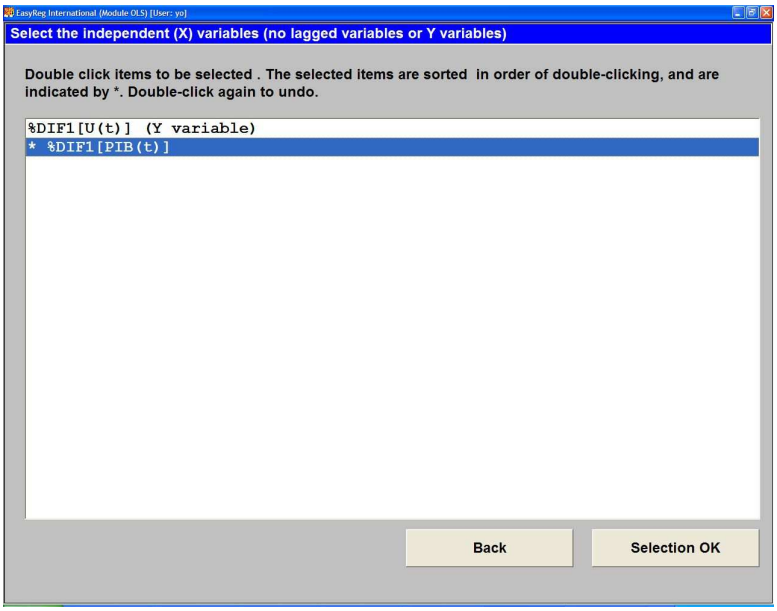
La siguiente ventana nos pide escoger las variables que entrarán en nuestro modelo. Escoja las variables " %DIF1[U(t)]", y " %DIF1[PIB(t)]" haciendo clic sobre cada una de las variables.



Haga clic en "Selection OK". La siguiente ventana nos pregunta si deseamos emplear una sub-muestra. En este caso queremos usar todos los datos, entonces haga clic en el botón "No" y después en "Continue".

La siguiente ventana nos pregunta por la variable dependiente (Y) de nuestro modelo. Escoja la variable " %DIF1[U(t)]" haciendo doble clic en dicha variable. Haga clic en "Continue" y nuevamente haga clic en el botón "Continue".

La siguiente ventana nos pregunta por las variables independientes de nuestro modelo (las X´s). Las variables independientes ya están preseleccionadas, revise que estén seleccionadas las variables independientes adecuadas. Haga clic en el botón "Selection OK".

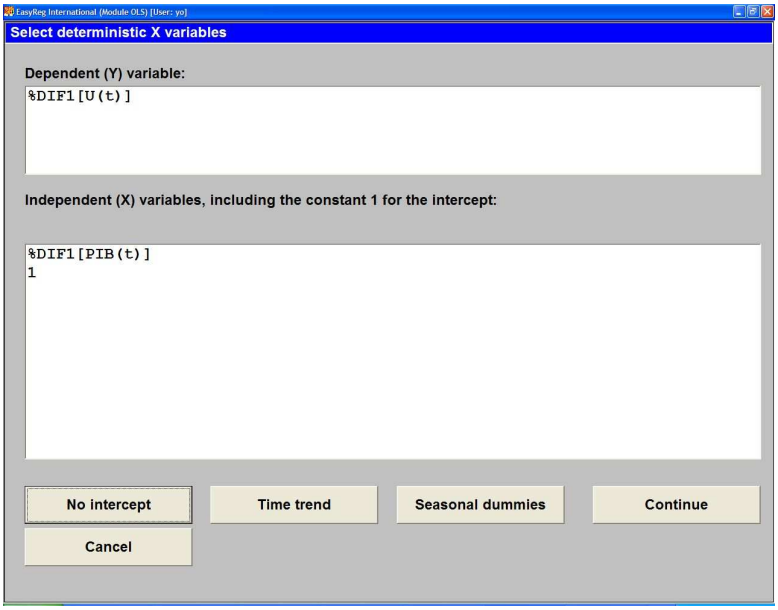


En la ventana siguiente se presenta un resumen de las variables, seleccione "continue" en esta ventana

y en la que se presenta a continuación "Selection ok". La siguiente ventana nos pregunta si queremos incluir rezagos de la variable dependiente como de variables explicativas, es decir Y_{t-1},Y_{t-2}, etc. En este caso no queremos incluir rezagos, haga clic en "Skip all".

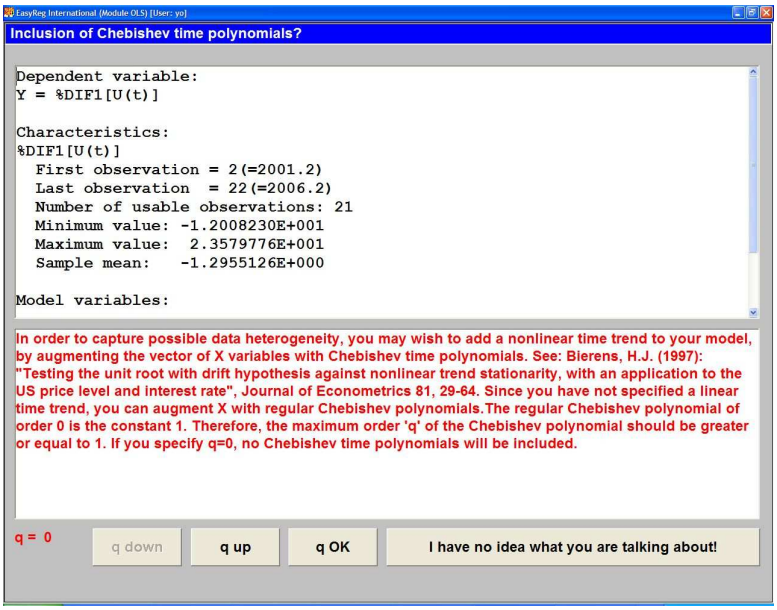
En la siguiente ventana nos pregunta si queremos incluir rezagos de las variables independientes. Haga doble clic únicamente en la variable " %DIF1[PIB(t)]" y luego clic en el botón "Selection OK".

Después de los anteriores pasos deben observar la siguiente ventana:



El botón "No intercept", en caso de ser presionado, quitará el intercepto del modelo. El botón "Time trend", en caso de ser hundido, incluirá una tendencia en la regresión, es decir el modelo estimado sería $\Delta\,\%U_t = \beta_0 + \beta_1 \Delta\,\%PIB_t + \beta_3 t + u_t$ para $t = 1..,22$ En este caso queremos tener un intercepto y no incluir una tendencia. Entonces haga clic en el botón "Continue".

La siguiente ventana nos pregunta si queremos incluir un polinomio del tiempo; es decir, el modelo quedaría $\Delta\,\%U_t = \beta_0 + \beta_1 \Delta\,\%PIB_t + \theta\,(t,q) + \varepsilon_t$ para $t = 1..,22$. Dónde $\theta\,(t,q)$ corresponde a un polinomio de Chevichev de orden q que depende del tiempo y que implicará una tendencia no lineal. Nosotros queremos tener un polinomio, lo que equivale a (esta es la opción predeterminada). Entonces haga clic en "q OK" y nuevamente en el botón "Continue" tres veces seguidas.



La siguiente ventana tiene los resultados de la estimación del modelo por el método de mínimos cuadrados ordinarios. Los resultados son reportados en el Recuadro 1.4

Recuadro 1.4 Resultados de la estimación por MCO del modelo 1.22

Dependent variable:
Y = %DIF1[U(t)]

Characteristics:
%DIF1[U(t)]
First observation = 2(=2001.02)
Last observation = 22(=2002.10)
Number of usable observations: 21
Minimum value: -1.2008230E+001
Maximum value: 2.3579776E+001
Sample mean: -1.2955126E+000

X variables:
X(1) = %DIF1[PIB(t)]
X(2) = 1

WARNING: The effective degrees or freedom is only 19.
Therefore, the estimation results may be unreliable!

Model:
Y = b(1)X(1) + b(2)X(2) + U,
where U is the error term, satisfying
E[U|X(1),X(2)] = 0.

OLS estimation results

Parameters	Estimate	t-value (S.E.) [p-value]	H.C. t-value (H.C. S.E.) [H.C. p-value]
b(1)	-1.4535084	-0.594 (2.44889) [0.55282]	-0.558 (2.60504) [0.57687]
b(2)	0.2408061	0.069 (3.47373) [0.94997]	0.063 (3.83814)
	[0.94473]		

Notes:
1: S.E. = Standard error
2: H.C. = Heteroskedasticity Consistent. These t-values and
standard errors are based on White's heteroskedasticity
consistent variance matrix.
3: The two-sided p-values are based on the normal approximation.

Effective sample size (n): 21
Variance of the residuals: 112.70520443
Standard error of the residuals (SER): 10.61627074
Residual sum of squares (RSS): 2141.39888425
(Also called SSR = Sum of Squared Residuals)
Total sum of squares (TSS): 2181.10332129
R-square: 0.0182
Adjusted R-square: -0.0335
Test for first-order autocorrelation:

Durbin-Watson test = 2.737553
REMARK: A better way of testing for autocorrelation
is to specify AR errors and then test the null
hypothesis that the AR parameters are zero.

Jarque-Bera/Salmon-Kiefer test = 5.064838
Null hypothesis: The errors are normally distributed
Null distribution: Chi-square(2))
p-value = 0.07947

Significance levels:	10 %	5 %
Critical values:	4.61	5.99
Conclusions:	reject	accept

Breusch-Pagan test = 0.232044
Null hypothesis: The errors are homoskedastic
Null distribution: Chi-square(1)
p-value = 0.63001

Significance levels:	10 %	5 %
Critical values:	2.71	3.84
Conclusions:	accept	accept
Information criteria:		
Akaike:	4.815168331	
Hannan-Quinn:	4.836757675	
Schwarz:	4.914646659	

En este caso, se necesita la matriz estimada de varianzas y covarianzas de los β_s' estimados expresada en la ecuación 1.20; en resumen se tiene que:

$$\begin{aligned}\hat{\beta} &= \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = \begin{bmatrix} 0{,}2408061 \\ -1{,}4535084 \end{bmatrix} \\ \sqrt{\widehat{Var(\beta_0)}} &= S_{\hat{\beta}_0} = 3{,}47373 \\ \sqrt{\widehat{Var(\beta_1)}} &= S_{\hat{\beta}_1} = 2{,}44889 \end{aligned} \tag{1.23}$$

Los resultados de la estimación se presentan normalmente en una tabla con un formato como el siguiente:

Cuadro 1.1: Estimación del modelo 1.22

	Variable dependiente $\Delta\,\%U_t$ Estadísticos t entre paréntesis
	Ecuación 1.22 2001:1 - 2006:2 MCO
Constante	0.2408061 (0.069)
$\Delta\,\%PIB_t$	-1.4535084 (-0.594)
R^2	0.0182
$R^2 ajustado$	-0.0335
No. de obs.	21

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos Cuadrados Ordinarios

Como se discutirá en el próximo capítulo, del cuadro 1.1 podemos concluir que el coeficiente asociado a la variación del PIB no es significativo ya que el estadístico t es relativamente pequeño (y el p-valor asociado a este es muy grande). Así, este coeficiente no es estadísticamente diferente de cero. Por lo tanto podemos concluir que la Ley de Okún no se cumple para Colombia durante el periodo estudiado, de acuerdo a la especificación empleada, es decir, los cambios en la tasa de crecimiento del PIB no afectan de ninguna forma la variación en la tasa de desempleo.

1.4. Ejercicios

El gobierno de una pequeña República está reconsiderando la viabilidad del transporte ferroviario, para lo cual contrata un estudio que determine un modelo que permita comprender de una forma más precisa el comportamiento de los ingresos del sector (I medidos en millones de dólares). Un investigador sugiere el siguiente modelo:

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 LDies_t + \alpha_5 LEI_t + \alpha_6 V_t + \varepsilon_t \tag{1.24}$$

Donde, CE_t, CD_t, $LDies_t$, LEI_t y V_t representan el consumo de electricidad medido en millones de Kilovatios/hora, el consumo de diesel medido en millones de galones, el número de locomotoras diesel en el país, el número de locomotoras eléctricas y el número de viajeros (medido en miles de pasajeros) en el año t, respectivamente.

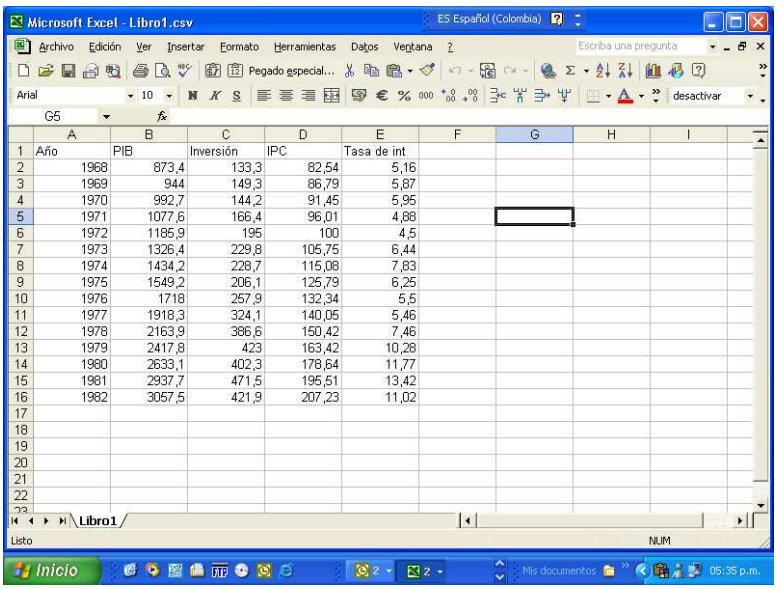
Para efectuar este estudio se cuenta con información para el período 1994-2003 (los datos se encuentran en el archivo regmult.xls).

1. De acuerdo con el enunciado anterior, estime el modelo 1.24 y reporte sus resultados en una tabla
2. Interprete los coeficientes estimados

1.5. Apéndice

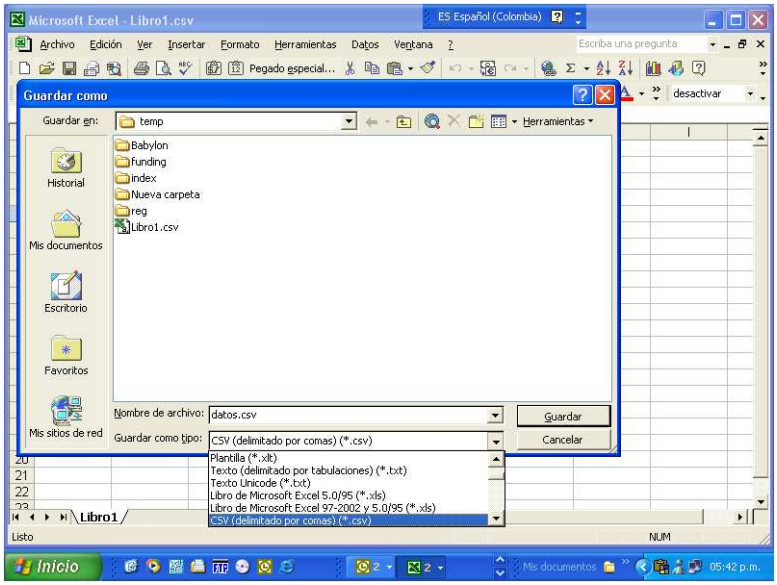
Apéndice 1.1 Guardando archivos en formato csv en Excel

Este apéndice describe los pasos necesarios para guardar un archivo de Excel en el formato de lectura requerido por EasyReg (csv). EasyReg lee las variables por columnas, de tal manera que la primera fila corresponderá a los nombres de las variables que deben ir entre comillas y separados por punto y coma. La segunda fila contendrá los datos correspondientes a la primera observación de cada una de las variables. Estos datos deben estar separados por punto y coma y no van entre comillas. Al final de cada línea no debe existir ningún signo. En otras palabras el formato debe ser el siguiente:
"Nombre Variable X_1";"Nombre Variable X_2";$\cdots$;"Nombre Variable X_k"
$X_{1,1}$;$X_{2,1}$;$\cdots$;$X_{k,1}$
$X_{1,2}$;$X_{2,2}$;$\cdots$;$X_{k,2}$
.....
$X_{1,n}$;$X_{2,n}$;$\cdots$;$X_{k,n}$
donde $X_{i,j}$ corresponde a la observación j de la variable X_i. Para esto prepare los datos en una hoja Excel de tal forma que cada una de las variables esté en una columna y la primera fila corresponde al nombre de cada variable. Es importante que los nombres de las variables estén en la primera fila de la hoja de Excel y adicionalmente la primera columna debe corresponder a la primera variable.



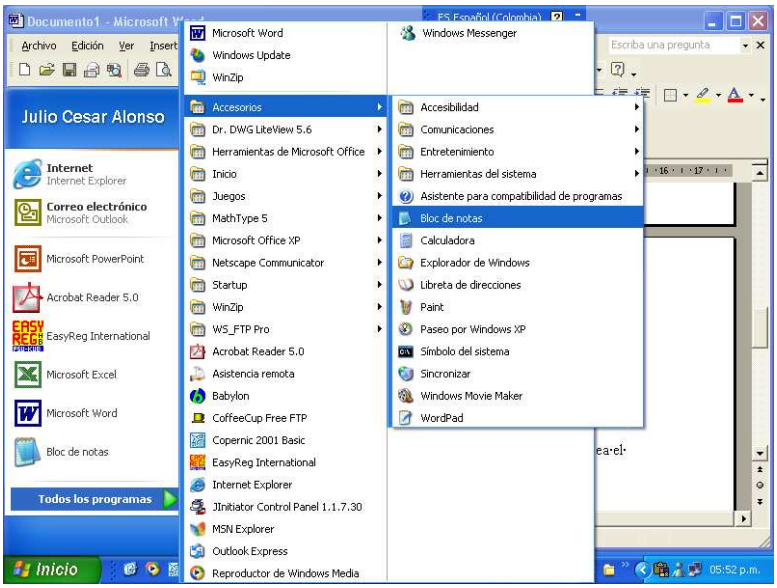
	A	B	C	D	E
1	Año	PIB	Inversión	IPC	Tasa de int
2	1968	873,4	133,3	82,54	5,16
3	1969	944	149,3	86,79	5,87
4	1970	992,7	144,2	91,45	5,95
5	1971	1077,6	166,4	96,01	4,88
6	1972	1185,9	195	100	4,5
7	1973	1326,4	229,8	105,75	6,44
8	1974	1434,2	228,7	115,08	7,83
9	1975	1549,2	206,1	125,79	6,25
10	1976	1718	257,9	132,34	5,5
11	1977	1918,3	324,1	140,05	5,46
12	1978	2163,9	386,6	150,42	7,46
13	1979	2417,8	423	163,42	10,28
14	1980	2633,1	402,3	178,64	11,77
15	1981	2937,7	471,5	195,51	13,42
16	1982	3057,5	421,9	207,23	11,02

Además, es muy importante que los números correspondientes a los datos no contengan ningún tipo de formato de Excel; por ejemplo, no queremos tener datos como 22.223,4 o 12.3 %. En caso que su hoja contenga datos con algún tipo de formato, deberá quitar el formato seleccionando en Excel todos los datos y haciendo clic en "Formato / Celdas" y escogiendo posteriormente la pestaña correspondiente a "número" y escoja la opción "General". Una vez los datos estén organizados, podemos grabarlos en el formato csv. Haga clic en "Archivo / Guardar como". En la ventana de diálogo que le aparecerá, identifique el fólder donde quiere guardar los datos.



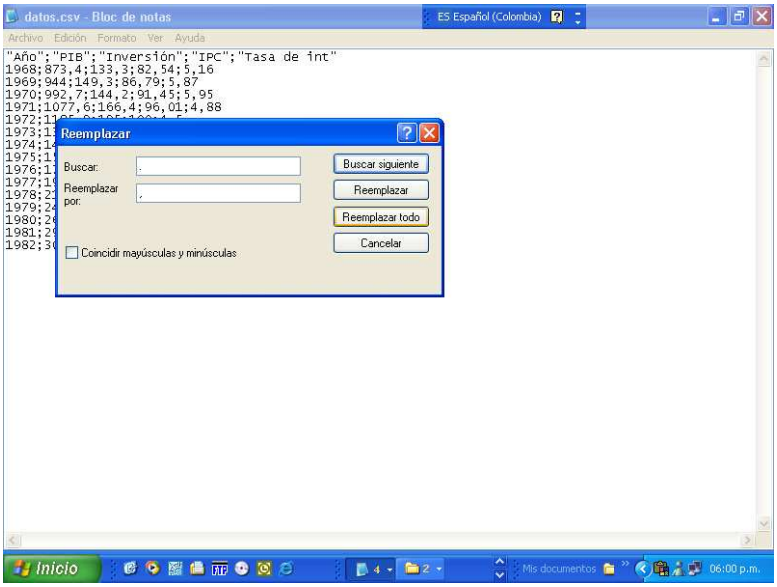
Asigne el nombre deseado para el archivo a crear en el espacio en blanco, al lado de "Nombre de Archivo". Haga clic en el triángulo que está al lado del botón cancelar. Esto abrirá un menú donde están los diferentes formatos en que puede grabar el archivo. Busque la opción "CSV (MS-DOS) (*.CSV)", y escójala haciendo clic sobre ella. Haga clic en el botón "Guardar". Los datos ya han sido grabados en el formato deseado.

Es posible que algunas versiones de Excel no guarden los datos exactamente en el formato requerido por EasyReg, por tanto necesitamos comprobar que el formato sea el correcto. Para esto, abra el Bloc de notas ("Inicio / Todos los Programas / Accesorios / bloc de notas").

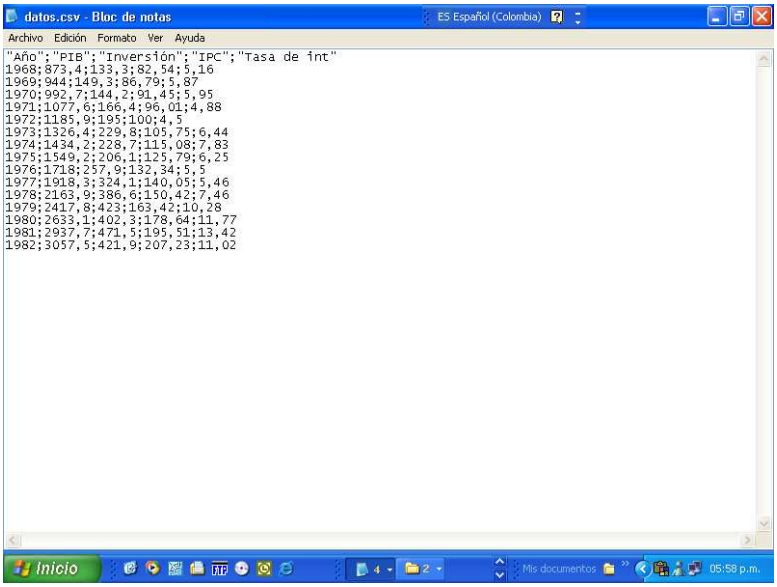


Abra el archivo csv que creó anteriormente. Verifique que los nombres de las variables estén entre comillas. En caso de que no lo estén, coloque estos nombres entre comillas. Además, verifique que los decimales estén denotados con comas y no con puntos. En caso de que los decimales estén escritos con puntos y no con comas, puede cambiar todos los puntos por comas usando "Edición / Remplazar" (ver

figura). Haga los cambios requeridos y guarde el archivo. (Edición/guardar)



Finalmente usted deberá tener un archivo que se vea como en la siguiente gráfica.



Apéndice 1.2 Derivación de los estimadores MCO

Formalmente, el estimador de MCO para el vector de coeficientes β en el modelo 1.14, denotado por $\hat{\beta}$, se encuentra minimizando la distancia cuadrada entre cada valor observado del vector de realizaciones de la variable aleatoria dependiente ($\mathbf{y}$) y el vector de estimaciones del modelo ($\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$). Formalmente, el problema es:

$$\underset{\hat{\beta}}{Min} \left\{ \left[\mathbf{y} - \mathbf{X}\hat{\beta} \right]^T \left[\mathbf{y} - \mathbf{X}\hat{\beta} \right] \right\} \tag{1.25}$$

Antes de encontrar las condiciones de primer orden y las de segundo orden para un mínimo, podemos simplificar un poco el problema expresado en la ecuación 1.25. Así, tenemos que:

$$\underset{\hat{\beta}}{Min} \left\{ \left[\mathbf{y^T} - \hat{\beta}^T \mathbf{X^T} \right] \left[\mathbf{y} - \mathbf{X}\hat{\beta} \right] \right\}$$

Al multiplicar los elementos dentro de los corchetes cuadrados obtenemos:

$$\underset{\hat{\beta}}{Min} \left\{ \mathbf{y^T y} - \hat{\beta}^T \mathbf{X^T y} - \mathbf{y^T X}\hat{\beta} + \hat{\beta}^T \mathbf{X^T X}\hat{\beta} \right\}$$

Observen que $\mathbf{y^T}_{\mathbf{1\times n}} \mathbf{X}_{\mathbf{n\times k}} \hat{\beta}_{k\times 1} = \left(\hat{\beta}^T_{k\times 1} \mathbf{X^T}_{\mathbf{n\times k}} \mathbf{y}_{\mathbf{1\times n}} \right)^T$ y además los productos $\hat{\beta}^T_{1\times k} \mathbf{X^T}_{\mathbf{k\times n}} \mathbf{y}_{\mathbf{n\times 1}}$ y $\mathbf{y^T}_{\mathbf{1\times n}} \mathbf{X}_{\mathbf{n\times k}} \hat{\beta}_{k\times 1}$ son escalares, por tanto $\mathbf{y^T}_{\mathbf{1\times n}} \mathbf{X}_{\mathbf{n\times k}} \hat{\beta}_{k\times 1} = \hat{\beta}^T_{1\times k} \mathbf{X^T}_{\mathbf{k\times n}} \mathbf{y}_{\mathbf{n\times 1}}$. Así, el problema 1.25 es equivalente a:

$$\underset{\hat{\beta}}{Min} \left\{ \mathbf{y^T y} - 2\hat{\beta}^T \mathbf{X^T y} + \hat{\beta}^T \mathbf{X^T X}\hat{\beta} \right\} \tag{1.26}$$

Ya podemos regresar a nuestro problema de minimización y considerar la condición de primer orden para este problema, en este caso la condición es:[13]

$$\frac{\partial}{\partial \hat{\beta}} \{\bullet\} = -2\mathbf{X^T y} + 2\mathbf{X^T X}\hat{\beta} \equiv 0 \tag{1.27}$$

La expresión 1.27 se conoce como las ecuaciones normales. Ahora, despejando $\hat{\beta}$, obtenemos:

$$2\mathbf{X^T X}\hat{\beta} = 2\mathbf{X^T y}$$

Multiplicando a ambos lados por la inversa[14] de $\mathbf{X^T X}$, tendremos:

$$\left(\mathbf{X^T X}\right)^{-1} \left(\mathbf{X^T X}\right) \hat{\beta} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T y}$$

Así, el estimado de MCO del vector de coeficientes β es:

$$\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T y}$$

La condición de segundo orden implica $\frac{\partial\{\bullet\}}{\partial\beta\,\partial\beta} = 2\mathbf{X^T X}$. Noten que $\left(\mathbf{X^T X}\right)$ es una matriz positiva semi-definida lo que garantiza que $\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T y}$ es un mínimo.

[13] Hay que anotar que la derivada es con respecto a un vector y no a un escalar.

[14] $\left(\mathbf{X^T X}\right)^{-1}$ existe pues $\mathbf{X}$ tiene rango completo gracias al supuesto de que $\mathbf{X_1}$,$\mathbf{X_2}$,...,$\mathbf{X_k}$ son linealmente independientes

Apéndice 1.3 Demostración del Teorema de Gauss-Markov

El Terorema de Gauss-Markov implica los siguientes supuestos:

- Existe una relación lineal entre $\mathbf{y}$ y las variables en la matriz $\mathbf{X}$ que se puede representar por el modelo lineal $\mathbf{y_{n\times 1}} = \mathbf{X_{n\times k}}\beta_{k\times 1} + \varepsilon_{\mathbf{n\times 1}}$
- $X_2, X_3, ..., X_k$ son no estocásticas y linealmente independientes (es decir $\mathbf{X}$ tiene rango completo y es una matriz no estocástica)
- El vector de errores ε tiene media cero, varianza constante y no autocorrelación. Es decir, $E[\varepsilon] = 0$ y $Var[\varepsilon] = \sigma^2\mathbf{I_n}$

Entonces el estimador de MCO, $\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$, es insesgado y eficiente. En otras palabras, $\hat{\beta}$ es el *Mejor Estimador Lineal Insesgado* (MELI) para el vector de coeficientes poblacionales β.

A continuación demostraremos este Teorema. Primero, demostremos que $\hat{\beta}$ es un estimador insesgado del vector de coeficientes poblacionales β. Formalmente, tenemos que demostrar que $E\left[\hat{\beta}\right] = \beta$. Para lograr tal fin calculemos el valor esperado del estimador MCO:

$$E\left[\hat{\beta}\right] = E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}\right]$$

Empleando el supuesto de que las variables explicativas son no estocásticas tenemos que,

$$E\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}E[\mathbf{y}]$$

Y por tanto,

$$E\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}E[\mathbf{X}\beta + \varepsilon]$$

$$E\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta + \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}E[\varepsilon]$$

Recuerden que habíamos supuesto que $E[\varepsilon] = 0$. Esto implica que

$$E\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta = I \cdot \beta$$

$$E\left[\hat{\beta}\right] = \beta$$

Así, hemos demostrado que el estimador MCO es insesgado.

Antes de continuar con la demostración del Teorema de Gauss-Markov, encontremos la varianza del estimador de MCO. Es decir,

$$Var\left[\hat{\beta}\right] = Var\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}\right]$$

$$Var\left[\hat{\beta}\right] = \left(\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\right)Var[\mathbf{y}]\left(\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\right)^T$$

$$Var\left[\hat{\beta}\right] = \left(\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\right)Var[\mathbf{y}]\left(\mathbf{X}\left(\left(\mathbf{X^T X}\right)^{-1}\right)^T\right)$$

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}Var[\mathbf{y}]\,\mathbf{X}\left(\mathbf{X^T X}\right)^{-1}$$

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}Var[\mathbf{X}\beta + \varepsilon]\,\mathbf{X}\left(\mathbf{X^T X}\right)^{-1}$$

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} Var\left[\varepsilon\right]\mathbf{X}\left(\mathbf{X^T X}\right)^{-1}$$

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\sigma^2\mathbf{X}\left(\mathbf{X^T X}\right)^{-1}$$

$$Var\left[\hat{\beta}\right] = \sigma^2\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\left(\mathbf{X^T X}\right)^{-1}$$

$$Var\left[\hat{\beta}\right] = \sigma^2\left(\mathbf{X^T X}\right)^{-1}$$

Ahora retornemos al Teorema de Gauss-Markov, para demostrarlo es necesario probar que este estimador tiene la mínima varianza entre todos los posibles estimadores lineales insesgados de β.

Sin perder generalidad, consideremos otro estimador lineal cualquiera $\tilde{\beta} = \left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]\mathbf{y}$. Si $\tilde{\beta}$ es insesgado, se debe cumplir que $E\left[\tilde{\beta}\right] = \beta$. Es decir:

$$E\left[\tilde{\beta}\right] = \left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]E\left[\mathbf{y}\right]$$

$$E\left[\tilde{\beta}\right] = \left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]E\left[\mathbf{X}\beta + \varepsilon\right] = \left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]\left[\mathbf{X}\beta + 0\right]$$

$$E\left[\tilde{\beta}\right] = \left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta + C\mathbf{X}\beta\right]$$

$$E\left[\tilde{\beta}\right] = \beta + C\mathbf{X}\beta$$

Para que este estimador sea insesgado, tiene que cumplirse que $C\mathbf{X} = 0$.

Ahora, analicemos la varianza de este nuevo estimador lineal insesgado.

$$Var\left[\tilde{\beta}\right] = Var\left[\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]\mathbf{y}\right]$$

$$Var\left[\tilde{\beta}\right] = \left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]Var\left[\mathbf{y}\right]\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]^T$$

$$Var\left[\tilde{\beta}\right] = \left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]\sigma^2 I_n\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]^T$$

$$Var\left[\tilde{\beta}\right] = \sigma^2\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} + C\right]\left[\mathbf{X}\left(\mathbf{X^T X}\right)^{-1} + C^T\right]$$

$$Var\left[\tilde{\beta}\right] = \sigma^2\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\left(\mathbf{X^T X}\right)^{-1} + C\mathbf{X}\left(\mathbf{X^T X}\right)^{-1} + \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}C^T + CC^T\right]$$

$$Var\left[\tilde{\beta}\right] = \sigma^2\left[\left(\mathbf{X^T X}\right)^{-1} + C\mathbf{X}\left(\mathbf{X^T X}\right)^{-1} + \left(\mathbf{X^T X}\right)^{-1}(C\mathbf{X})^T + CC^T\right]$$

Como la condición $C\mathbf{X} = 0$ para que $\tilde{\beta}$ sea insesgado tiene que cumplirse, entonces:

$$Var\left[\tilde{\beta}\right] = \sigma^2\left[\left(\mathbf{X^T X}\right)^{-1} + CC^T\right]$$

CC^T es una matriz cuyos elementos en la diagonal principal serán positivos.[15] (¿Por qué?) Ahora comparemos la varianza de nuestro estimador de MCO ($\hat{\beta}$) con $\tilde{\beta}$. Recuerden que $Var\left[\hat{\beta}\right] = \sigma^2\left(\mathbf{X^T X}\right)^{-1}$, adicionalmente observen que $Var\left[\tilde{\beta}\right] = \sigma^2\left[\left(\mathbf{X^T X}\right)^{-1} + CC^T\right]$ no puede ser menor que $\sigma^2\left(\mathbf{X^T X}\right)^{-1}$, pues:

[15] Una matriz positiva semi-definida

- $Var\left[\tilde{\beta}_i\right] = \sigma^2\left\{\left[(\mathbf{XX})^{-1}_{ii} + CC^T_{ii}\right]\right\}$, donde A_{ij} corresponde al elemento en la fila i y columna j de la matriz A, y
- CC^T_{ii} es positivo.

Por tanto $Var\left[\tilde{\beta}_i\right] = \sigma^2\left\{\left[(\mathbf{XX})^{-1}_{ii} + CC^T_{ii}\right]\right\} > Var\left[\hat{\beta}_i\right] = \sigma^2\left[(\mathbf{XX})^{-1}_{ii}\right]$. En el mejor de los casos $Var\left[\tilde{\beta}\right] = \sigma^2\left(\mathbf{X^T X}\right)^{-1}$ y eso sólo ocurre cuando $C = 0$. En este caso $\tilde{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y} = \hat{\beta}$, es decir la mínima varianza posible de un estimador lineal insesgado es igual a la varianza del estimador MCO. Por tanto $\hat{\beta}$ es MELI.

Apéndice 1.4 Algunas propiedades importantes de los residuos estimados.
Como se discutió anteriormente, se tiene que los residuos estimados están definidos como

$$\hat{\varepsilon} = y - \mathbf{X}\hat{\beta}$$

La primera propiedad que cumple el vector de residuos ($\hat{\varepsilon}$) es:

$$\mathbf{X}^T\hat{\varepsilon} = 0 \tag{1.28}$$

Para demostrar esta propiedad podemos partir de la definición de los residuos estimados. Es decir, tenemos que:

$$\mathbf{X}^T\hat{\varepsilon} = \mathbf{X}^T\left[y - \mathbf{X}\hat{\beta}\right] = \mathbf{X}^T y - \mathbf{X}^T\mathbf{X}\hat{\beta}$$

Además, sustituyendo $\hat{\beta}$ se tiene

$$\mathbf{X}^T\hat{\varepsilon} = \mathbf{X}^T y - \mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y$$

Por lo tanto,

$$\mathbf{X}^T\hat{\varepsilon} = \mathbf{X}^T y - \mathbf{X}^T y = 0$$

De esta propiedad se desprende un resultado muy interesante. Dado que esta propiedad implica que:

$$\mathbf{X}^T\hat{\varepsilon} = \begin{bmatrix} \sum_{i=1}^{n} X_{1i}\hat{\varepsilon}_i \\ \sum_{i=1}^{n} X_{2i}\hat{\varepsilon}_i \\ \vdots \\ \sum_{i=1}^{n} X_{ki}\hat{\varepsilon}_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Entonces, se desprende un resultado importante. *Si el modelo tiene intercepto la suma de los residuos estimados será cero.* Esta afirmación se puede demostrar facilmente. Noten que en caso de tener intercepto el modelo, $\mathbf{X}$ tendrá una columna de unos. Por ejemplo, $X_{1i} = 1$ y se desprende que:

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0 \tag{1.29}$$

Este último resultado implica que *la media del residuo estimado sea cero.* En otras palabras:

$$\bar{\hat{\varepsilon}} = \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i}{n} = 0 \tag{1.30}$$

La segunda propiedad de los residuos estimados que discutiremos es que:

$$E\left[\hat{\varepsilon}^T\hat{\varepsilon}\right] = E\left[\sum_{i=1}^{n} \hat{\varepsilon}_i^2\right] = (n-2)\,\sigma^2 \tag{1.31}$$

Para demostrar esta propiedad, reescribamos de manera más conveniente los residuos estimados:

$$\hat{\varepsilon} = y - \mathbf{X}\hat{\beta} = y - \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T y$$

$$\hat{\varepsilon} = \left[I - \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \right] y$$

Ahora, definamos la matriz $\mathbf{M} = I - \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T$. Es muy facil mostrar (hágalo) que $\mathbf{M}$ es idempotente y simétrica. Entonces, se tiene que los residuos estimados se pueden expresar de la siguiente manera:

$$\hat{\varepsilon} = \mathbf{M}y = \mathbf{M}\mathbf{X}\beta + \mathbf{M}\varepsilon = \mathbf{M}\varepsilon \tag{1.32}$$

Es importante anotar que $\mathbf{M}\mathbf{X}\beta = 0$, dado que

$$\mathbf{M}\mathbf{X}\beta = \left[I - \mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \right] \mathbf{X}\beta = [\mathbf{X}\beta - \mathbf{X}\beta] = 0$$

Regresando a (1.32), se tiene que:

$$E\left[\hat{\varepsilon}^T \hat{\varepsilon} \right] = E\left[(\mathbf{M}\varepsilon)^T \mathbf{M}\varepsilon \right] = E\left[\varepsilon^T \mathbf{M}^T \mathbf{M}\varepsilon \right] = E\left[\varepsilon^T \mathbf{M}\varepsilon \right]$$

Y dado que el producto $\hat{\varepsilon}^T \hat{\varepsilon}$ es un escalar tendremos que:

$$E\left[\hat{\varepsilon}^T \hat{\varepsilon} \right] = trace\left(\varepsilon^T \mathbf{M}\varepsilon \right) = trace\left(\mathbf{M}\varepsilon^T \varepsilon \right)$$

Empleando las propiedades de la traza,

$$E\left[\hat{\varepsilon}^T \hat{\varepsilon} \right] = trace\left(\mathbf{M}\sigma^2 I \right) = \sigma^2 trace\left(\mathbf{M} \right)$$

Noten que encontrar la traza de la matriz $\mathbf{M}$ es muy sencillo pues,

$$trace\left(\mathbf{M} \right) = trace\left(I_n \right) - trace\left(\mathbf{X} \left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \right)$$

y empleando otras propiedades de la traza sabemos que:

$$trace\left(\mathbf{M} \right) = trace\left(I_n \right) - trace\left(\left(\mathbf{X}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{X} \right)$$

De esta manera se tiene que

$$trace\left(\mathbf{M} \right) = trace\left(I_n \right) - trace\left(I_k \right) = n - k$$

Esto implica que

$$E\left[\hat{\varepsilon}^T \hat{\varepsilon} \right] = \sigma^2 trace\left(\mathbf{M} \right) = \left(n - k \right) \sigma^2 \tag{1.33}$$

Este último resultado es importante porque implica que *el estimador MCO para la varianza de los errores es insesgado*. En otras palabras:

$$E\left[s^2 \right] = E\left[\frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n - k} \right] = E\left[\frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n - k} \right] = \sigma^2 \tag{1.34}$$

CAPÍTULO 2

Inferencia y análisis de regresión

Objetivos del capítulo

Al finalizar este capítulo, el lector estará en capacidad de:

- Realizar por medio de EasyReg pruebas individuales y conjuntas sobre los coeficientes estimados de un modelo de regresión.
- Valorar mediante la salida de EasyReg la bondad estadística del modelo estimado.
- Identificar en la salida de EasyReg las diferentes fuentes de variación de la variable dependiente.
- Construir la tabla Anova a partir de los datos que se encuentran en la salida de las estimaciones de los diferentes modelos estimados mediante EasyReg.

2.1. Introducción

El término inferencia se refiere a verificar si ciertas restricciones que impone la teoría económica son válidas o no, para la muestra a partir de la cual hemos estimado nuestro modelo. En este capítulo estudiaremos cómo comprobar restricciones que involucren uno o más parámetros del modelo estudiado. En el Capítulo 1 estudiamos el método de Mínimos Cuadrados Ordinarios (MCO) para estimar un modelo lineal de la forma:

$$\mathbf{y_{n\times 1}} = \mathbf{X_{n\times k}}\beta_{k\times 1} + \varepsilon_{n\times 1} \tag{2.1}$$

Como lo discutimos, el estimador de MCO para el parámetro β está dado por:

$$\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y} \tag{2.2}$$

Adicionalmente, demostramos que si se supone que i) $X_2, X_3, ..., X_k$ son fijas y linealmente independientes, y ii) $E[\varepsilon] = 0$, $Var[\varepsilon] = \sigma^2\mathbf{I_n}$; entonces el estimador de MCO es el mejor estimador lineal insesgado.

Pero, la probabilidad de que el estimador de MCO ($\hat{\beta}$) para una muestra dada sea exactamente igual al valor poblacional es cero. Aunque esta última afirmación suene a primera vista algo contradictoria; si se reflexiona un poco tendrá más sentido. Es muy difícil "adivinar" correctamente el valor real a partir de un solo intento, cuando el verdadero valor pertenece a un conjunto infinito de valores.

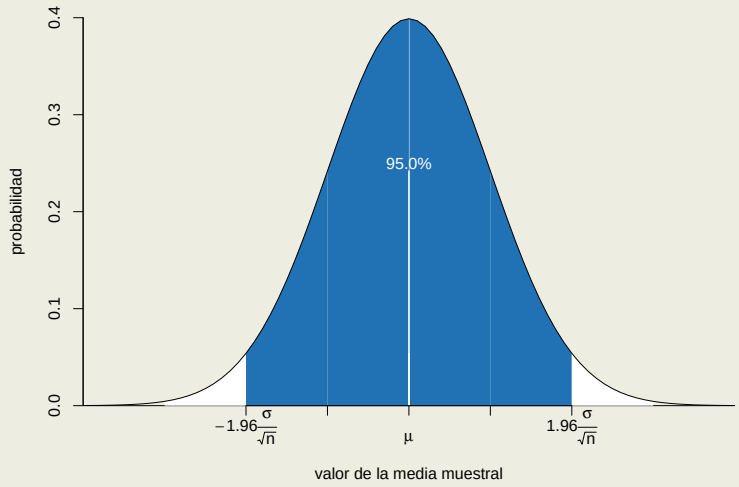
Ustedes se deben estar preguntando: si eso es cierto, entonces ¿para qué empleamos estos estimadores si dada una muestra no hay chance que el estimador sea igual al valor real? La respuesta es clara, como se demostró en el capítulo anterior, en promedio el estimador de MCO es igual al valor poblacional (insesgado), por tanto si recolectamos un número lo suficientemente grande de muestras podremos encontrar el valor promedio del estimador. Pero claramente en los estudios económicos no nos podemos dar el "lujo" de tener más de una muestra para la población en estudio.

Es ahí donde la teoría estadística llega al auxilio del investigador. Si bien sabemos que un valor estimado a partir de una muestra tiene una probabilidad de cero de acertar el valor real, si podemos aumentar la certidumbre de nuestra estimación si conocemos la dispersión y la distribución del estimador (Ver Ejemplo 2.3).

Ejemplo 2.1

Supongamos que ustedes quieren conocer el peso medio de una población de estudiantes, es decir la media poblacional del peso de ellos. En este caso podrían hacer un censo del peso de todos los estudiantes de la universidad y una vez pesados todos calcular el promedio aritmético. Pero, hacer un censo es muy costoso y dispendioso, por tanto, como aprendieron en sus cursos introductorios de estadística, se acostumbra seleccionar una muestra aleatoria de tamaño n y a cada uno de esos n estudiantes se les pesa (W_i para $i = 1, 2, ..., n$). A partir de esta muestra podemos calcular la media muestral, es decir $\bar{w} = \frac{\sum_{i=1}^{n} W_i}{n}$. ¿Cuál es la probabilidad que $\bar{w}$ sea exactamente igual a la media poblacional?. Claramente esta probabilidad es cero; pero gracias al Teorema del Límite Central, si la muestra es lo suficientemente grande, entonces sabemos que $\bar{w}$ sigue una distribución normal con una

media igual a la media poblacional y varianza igual a la varianza poblacional dividida por el tamaño muestral ($\frac{\sigma^2}{n}$). Este resultado nos permite aumentar la certidumbre de nuestro pronóstico ($\bar{w}$), para que sea más exacto



Sabemos que si nos movemos dos desviaciones estándar (en este caso $\frac{\sigma}{\sqrt{n}}$) a la derecha y a la izquierda de nuestro valor estimado, entonces tenemos un 95 % de seguridad de que el valor real está contenido en ese intervalo. Así, empleando la distribución de nuestro estimador podemos aumentar la certidumbre sobre nuestra estimación.

Si conocemos la distribución de nuestro estimador de MCO podremos mejorar la certeza en torno a nuestras estimaciones. Estudiemos pues cuál es la distribución de nuestro estimador.

Sabemos que $E\left[\hat{\beta}\right] = \beta$ y $Var\left[\hat{\beta}\right] = \sigma^2\left(\mathbf{X^T X}\right)^{-1}$. Por tanto, ahora necesitamos conocer la distribución exacta que sigue nuestro estimador de MCO, pues recuerden que una función de distribución no está caracterizada únicamente por su media y varianza. Por otro lado, $\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$ es una combinación lineal de variables no aleatorias y variables aleatorias. Noten que $\left(\mathbf{X^T X}\right)^{-1}$ corresponde a una matriz no aleatoria, pues hemos asumido que $\mathbf{X}$ es una matriz no estocástica. Además, dado que cada uno de los $\mathbf{y_i}$ es una variable aleatoria, pues $\mathbf{y_i}$ depende de ε_i; tenemos que $\mathbf{X^T y}$ es un vector cuyos elementos son sumas de variables aleatorias. Es decir,

$$\mathbf{X^T y} = \begin{bmatrix} \sum_{i=1}^{n} \mathbf{y_i} \\ \sum_{i=1}^{n} \mathbf{y_i X_{2i}} \\ \sum_{i=1}^{n} \mathbf{y_i X_{3i}} \\ \vdots \\ \sum_{i=1}^{n} \mathbf{y_i X_{ki}} \end{bmatrix}$$

Por lo tanto $\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$ también será un vector cuyos elementos corresponden a combinaciones

lineales de sumas de variables aleatorias. Ahora bien, si tenemos una muestra lo suficientemente grande, podemos emplear el Teorema del Límite Central para concluir que el vector $\hat{\beta}$ al ser una combinación de suma de variables aleatorias seguirá una distribución normal.

Así, si hay una muestra lo suficientemente grande tendremos que:

$$\hat{\beta} \sim N_k \left(\beta, \left[\sigma^2 \left(\mathbf{X^T X} \right)^{-1} \right]_{k \times k} \right) \tag{2.3}$$

En otras palabras, los estimadores de MCO seguirán una distribución asintóticamente normal (sigue una distribución Multivariada Normal de orden k) con media igual al valor poblacional y matriz de varianzas y covarianzas $\sigma^2 \left(\mathbf{X^T X} \right)^{-1}$.

De esta manera, cualquier estimador $\hat{\beta}_p$, para $1 \leq p \leq k$, seguirá una distribución asintótica normal con media igual al valor poblacional β_p y varianza igual al elemento en la columna y fila p de la matriz de varianzas y covarianzas de los estimadores de MCO. En otras palabras,

$$\hat{\beta}_p \sim N \left(\beta_p, \left[\sigma^2 \left\{ Elemento\ de\ la\ matriz \left(\mathbf{X^T X} \right)^{-1} \right\} \right] \right) \tag{2.4}$$

Una vez conocemos la distribución del estimador de MCO podremos hacer inferencia sobre los parámetros poblacionales. En lo que resta de este capítulo discutiremos cómo hacer inferencia respecto a un parámetro (pruebas individuales) y sobre un conjunto de ellos (pruebas conjuntas).

Noten que este resultado será cierto únicamente si la muestra es grande. En el caso que no lo sea, entonces se necesitará suponer que el término aleatorio del error sigue una distribución normal. En el apéndice 2.2 se presenta una prueba que permite comprobar si los residuos estimados provienen o no de una población que sigue una distribución normal.

2.2. Pruebas individuales sobre los parámetros

Hay un pequeño inconveniente al emplear el resultado de la ecuación 2.4. En la práctica, no conocemos el valor poblacional de la varianza del error σ^2, por tanto debemos estimarlo por medio de s^2. Al estimar este parámetro, estamos introduciendo más incertidumbre en nuestro proceso de inferencia. Esta incertidumbre, agregada al problema de inferencia, implica que la distribución de los estimadores no será tan concentrada hacia la media, como lo es una distribución normal cuyas colas son relativamente pequeñas. La mayor incertidumbre implicará una mayor probabilidad de estar lejos del valor de la media poblacional, que se refleja en unas colas más grandes (más altas). Estas colas más altas, aun manteniendo la simetría y gran masa de probabilidad cercana a la media, se pueden capturar con una distribución t.

Por tanto, cuando no conocemos la varianza poblacional del término aleatorio de error, cada uno de los estimadores de MCO seguirá una distribución t, con $n - k$ grados de libertad. De esta manera, tenemos que un *intervalo de confianza* para β_k del $(1 - \alpha)\,100\,\%$ de confianza está dado por:

$$\hat{\beta}_p \pm t_{\frac{\alpha}{2}, n-k} s_{\hat{\beta}_p} \tag{2.5}$$

donde $s_{\hat{\beta}_p}$ corresponde a la raíz cuadrada de la varianza estimada del estimador $\hat{\beta}_k$.

Además, la distribución del estimador nos permite probar la hipótesis nula que el valor poblacional β_p es igual a cualquier constante c, es decir $H_o : \beta_p = c$; versus la hipótesis alterna que β_p no es igual

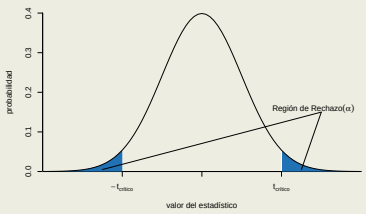
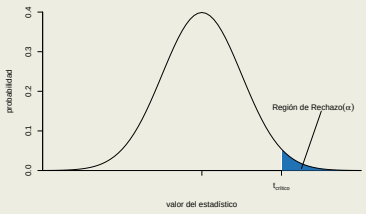
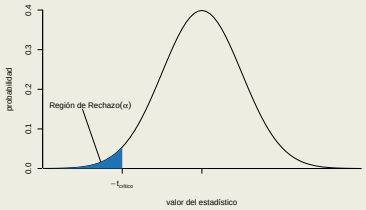
a la constante c, $H_A : \beta_p \neq c$. Para contrastar estas hipótesis, se emplea el siguiente estadístico t:

$$t_c = \frac{\hat{\beta}_p - c}{s_{\hat{\beta}_p}} \tag{2.6}$$

Este estadístico t_c (o t-calculado) sigue una distribución t con $n-k$ grados de libertad; por tanto, se rechazará la hipótesis nula si el valor absoluto del estadístico t_c es lo suficientemente grande. En otras palabras, si $|t_c| > t_{\frac{\alpha}{2},n-k}$.

Un caso muy especial es cuando $c = 0$, en ese caso la hipótesis nula implicará $\beta_k = 0$; es decir, la variable X_k no explica la variable dependiente. Por otro lado, la hipótesis alterna implicará que $\beta_k \neq 0$; es decir, la variable X_k sí ayuda a explicar la variable dependiente. Cuando se concluye a partir de la muestra que $\beta_k \neq 0$, entonces se dice que la variable X_k es significativa para el modelo, alternativamente se dice que el coeficiente β_k es significativo. Por esto, este tipo de pruebas se conocen como pruebas de significancia individual.

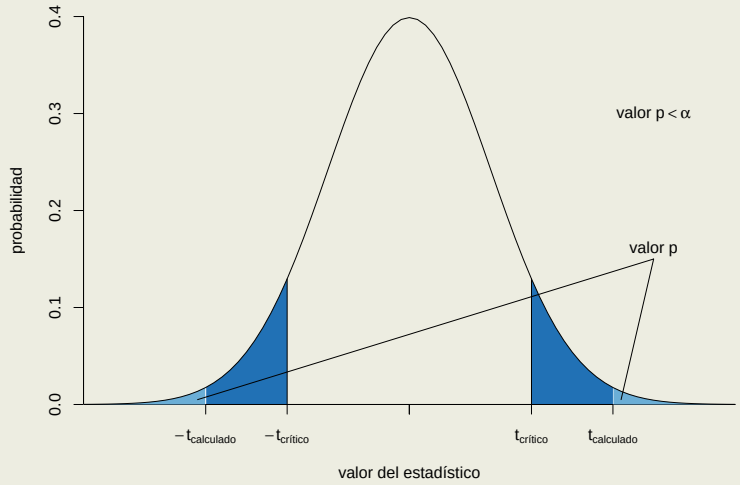
Ejemplo 2.2 Tipos de pruebas individuales

Hipótesis nula	**Hipótesis alterna**	**Estadístico**	**Rechazar H_0 si**	**Zona de rechazo**
$H_o : \beta_p = c$	$H_A : \beta_p \neq c$	$t_c = \frac{\hat{\beta}_p - c}{s_{\hat{\beta}_p}}$	$\lvert t_c \rvert > t_{\frac{\alpha}{2}}$	
$H_o : \beta_p \leq c$	$H_A : \beta_p > c$	$t_c = \frac{\hat{\beta}_p - c}{s_{\hat{\beta}_p}}$	$t_c > t_\alpha$	
$H_o : \beta_p \geq c$	$H_A : \beta_p < c$	$t_c = \frac{\hat{\beta}_p - c}{s_{\hat{\beta}_p}}$	$t_c < t_\alpha$	

Ejemplo 2.3 Pruebas de hipótesis y el valor p

Una manera alternativa de rechazar o no H_0 es empleando el valor p. Este valor corresponde a la probabilidad de obtener un estadístico t más grande en valor absoluto que el observado. Empleando este criterio, la decisión de rechazar se puede tomar de la siguiente manera:

Nivel de significancia	**Se rechaza si**
10 %	valor p $<$0.1
5 %	valor p $<$0.05
1 %	valor p $<$0.01



2.3. Descomposición de la variación de la variable dependiente

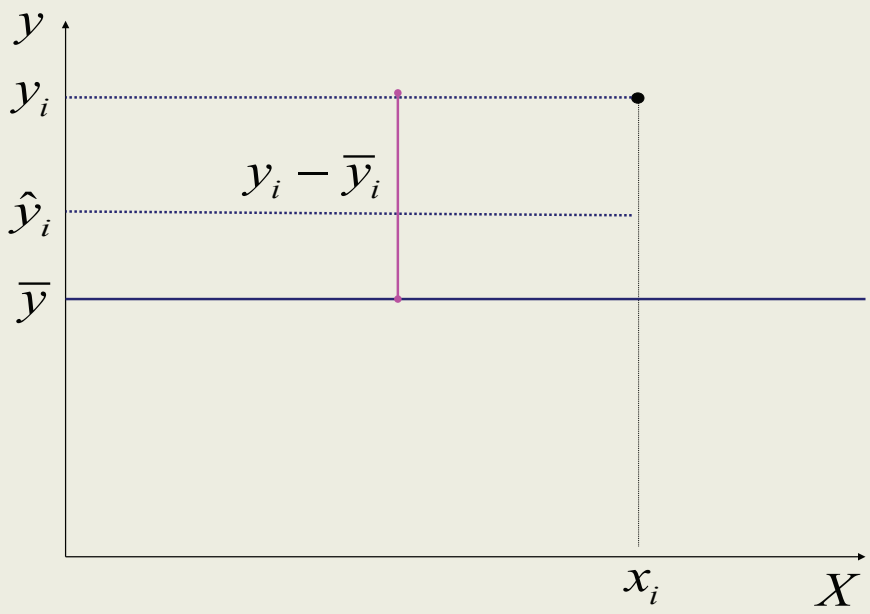
Antes de continuar, es necesario preguntarnos ¿cómo determinar si el modelo estimado explica satisfactoriamente la muestra bajo estudio?. En especial, ¿cómo podemos saber si el modelo lineal en efecto es válido para la muestra?, pues recuerde que nuestro primer supuesto es que la relación entre las variables independientes y la dependiente es lineal, supuesto simplificador que en la mayoría de los casos no es sugerido por la teoría económica.

Antes de analizar qué tan bien explica nuestro modelo la variable dependiente, es necesario conocer cómo se puede descomponer la variabilidad de la variable dependiente. Inicialmente consideremos la variación total de la variable dependiente con respecto a su media $(y_i - \bar{y})$, esta variación puede ser descompuesta en dos partes. La primera de ellas es la parte de la variación que es explicada por el modelo; esta parte de la variación es la diferencia que existe entre lo que predice el modelo para la

observación[1] i ($\hat{y}_i$) y nuestra mejor predicción si no contáramos con un modelo, es decir la media de la variable dependiente ($\bar{y}$). En otras palabras, la parte de la variación de la variable dependiente explicada por el modelo corresponde a $(\hat{y}_i - \bar{y})$. La segunda parte de la variación de la variable dependiente es la que no puede ser explicada por el modelo, que denominaremos el error o residuo, es decir $\hat{\epsilon}_i = (y_i - \hat{y}_i)$. Así, $(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$. Esta descomposición se puede entender de forma más intuitiva con el recuadro 2.1.

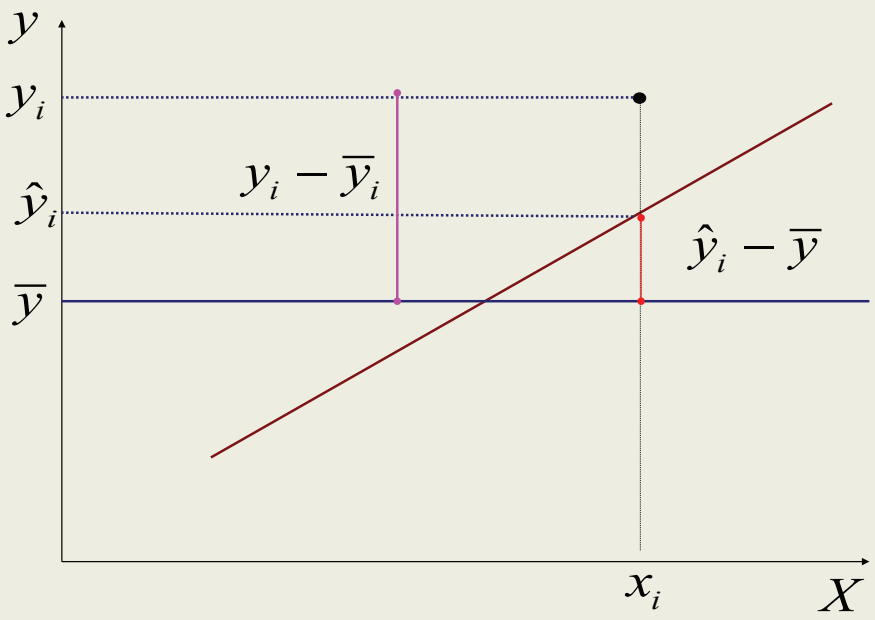
Recuadro 2.1 Descomposición de la variación total de la variable dependiente

Si no contamos con un modelo, la mejor forma de explicar la variable dependiente es empleando su media. A la diferencia entre la media y cada observación se le conoce como la variación total.

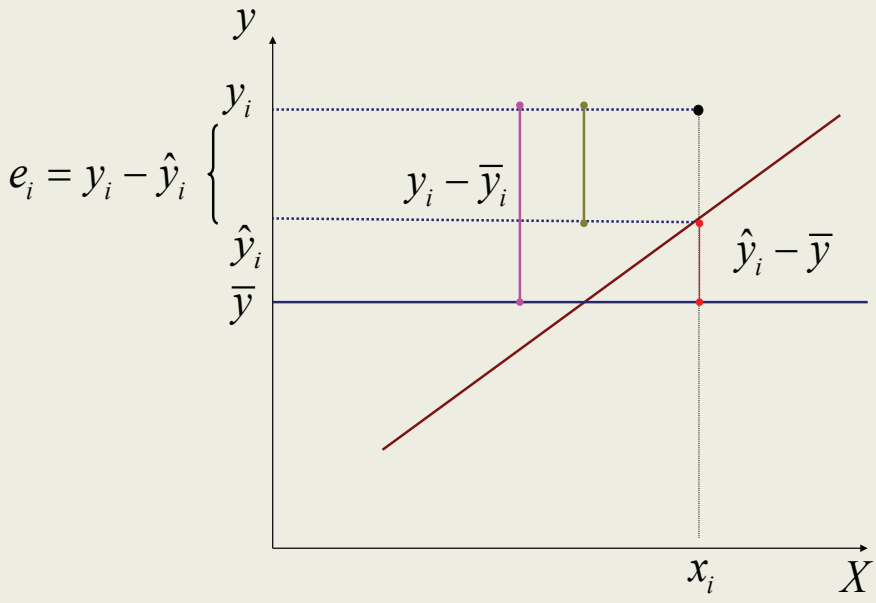


La variación total se puede descomponer en: i) la parte explicada por el modelo $(\hat{y}_i - \bar{y})$, y

[1] $\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki}$ para $i = 1, 2, ..., n$, o en forma matricial $\hat{y} = X\hat{\beta}$.



ii) la parte que el modelo no explica (el error) $\hat{\epsilon}_i = (y_i - \hat{y}_i)$



Si queremos conocer la variación total de la variable dependiente para toda la muestra, es buena idea considerar la suma de todas las variaciones al cuadrado para evitar el efecto contrario, de variaciones por encima y por debajo de la media. Llamaremos a $\sum_{i=1}^{n} (y_i - \bar{y})^2$ la Suma Total al Cuadrado (SST por su nombre en inglés: Sum of Squares Total). Es muy fácil mostrar que en presencia de un modelo

con intercepto:[2]

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2+\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \tag{2.7}$$

La parte de esta variación total al cuadrado que es explicada por el modelo será $\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2$, la denotaremos como SSR (por su nombre en inglés: Sum of Squares of the Regression). Y la parte que no es explicada por la regresión se denotará por SSE (por su nombre en inglés: Sum of Squares Error). Estas sumas al cuadrado también se pueden expresar en forma matricial. Es relativamente fácil mostrar que:

$$SST=\sum_{i=1}^{n}(y_i-\hat{y})^2=\mathbf{y^T y}-n\bar{\mathbf{y}}^2$$

$$SSR=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2=\hat{\beta}^T\mathbf{X^T y}-n\bar{\mathbf{y}}^2$$

$$SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\mathbf{y^T y}-\hat{\beta}^T\mathbf{X^T y}$$

Así tendremos que

$$SST=SSR+SSE$$

Una clara medida de qué tan bueno es nuestro modelo es examinar qué porcentaje de la variación total es explicada por el modelo, a esto se le conoce como el R^2. Formalmente,

$$R^2=\frac{SSR}{SST}=1-\frac{SSE}{SST}$$

Dado que el R^2 es un porcentaje, éste estará entre cero y uno; es decir, $0\leq R^2\leq 1$. Si $R^2=1$, entonces toda la variación del modelo es explicada por el modelo; para el caso de dos variables explicativas, esto implicará que todos los puntos se encuentran sobre el plano estimado por el modelo, en caso de tener una sola variable explicativa, este R^2 implicará que todos los puntos están sobre la línea de regresión. Por otro lado, si $R^2=0$ tendremos que nuestro modelo no explica nada de la variación de la variable dependiente.

Una forma de resumir la descomposición de la variabilidad de la variable dependiente[3] es emplear una tabla conocida como la tabla ANOVA (Analysis of Variance.). La tabla ANOVA no sólo resume la descomposición de la variabilidad de la variable dependiente, sino que también resume información importante como los grados de libertad asociados con cada suma al cuadrado.

Intuitivamente, noten que para calcular el $SST=\sum_{i=1}^{n}(y_i-\bar{y})^2$ se emplean n observaciones y se pierde un grado de libertad al emplear la media, por tanto los grados de libertad del SST son $n-1$. Por otro lado, para calcular el $SSE=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ se emplean n observaciones y se pierde k grados de libertad al calcular $\hat{y}_i$, por tanto los grados de libertad del SSE son $n-k$. Finalmente, así como $SST=SSE+SSR$, también se debe cumplir que los grados de libertad del SST deben ser iguales a la suma de los grados de libertad del SSE y el SSR. Así por diferencia, podemos encontrar que los

[2]En el apéndice 1 se presenta una demostración de esta afirmación.

[3]Esta es una alternativa a la gráfica que no funcionaría para un modelo con más de dos variables explicativas. Aun en el caso de un modelo con una sola variable dependiente, el análisis gráfico se convierte en una situación inmanejable.

grados de libertad del SSR son $k-1$. Una forma alternativa para llegar a este último resultado es advertir que para calcular $SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$ se emplean k grados de libertad en el cálculo de cada $\hat{y}_i$ y se pierde un grado de libertad al emplear la media $\bar{y}_i$, por tanto los grados de libertad del SSR son $k-1$.

Finalmente, en la tabla ANOVA se reporta la Media Cuadrada (MS) de los errores y de la regresión que corresponde a las respectivas Sumas al Cuadrado divididas por sus respectivos grados de libertad. En especial, el MS correspondiente al error, MSE, es exactamente igual al estimador de la varianza del error. Claramente de la tabla ANOVA, también se puede derivar rápidamente el R^2, pues toda la información necesaria para el cálculo de este estadístico está disponible en la tabla.

Recuadro 2.2 Tabla ANOVA

Fuente de la Variación	**SS**	**Grados de libertad**	**MS**
Regresión	$SSR = \hat{\beta}^T \mathbf{X^T y} - n\bar{\mathbf{y}}^2$	$k-1$	$MSRN = \frac{SSRN}{k-1}$
Error	$SSE = \mathbf{y^T y} - \hat{\beta}^T \mathbf{X^T y}$	$n-k$	$MSR = s^2 = \frac{SSR}{n-k}$
Total	$SST = \mathbf{y^T y} - n\bar{\mathbf{y}}^2$	$n-1$	

Antes de continuar, es importante anotar algunas consideraciones para tener en cuenta sobre la medida de bondad de ajuste brindada por el R^2. Primero, es relativamente fácil mostrar que en caso de que el modelo estimado no contenga un intercepto, entonces $SST \neq SSRN + SSE$ (ver anexo 1). Por tanto en ese caso el R^2 no representará la parte de la variabilidad total de la variable dependiente explicada por el modelo, por eso se dice que el R^2 carece de interpretación cuando el modelo estimado no posee intercepto.

Segundo, dado que el R^2 es una medida del ajuste de un modelo lineal estimado a los datos, entonces el R^2 se puede emplear para escoger entre diferentes modelos lineales el que se ajuste mejor a los datos,pero si se cumplen algunas condiciones.

Por ejemplo, supongamos que se quiere estimar la función de demanda de un bien Q_i, para lo cual se cuenta con información de su precio, p_i, el precio de un bien sustituto, ps_i, y un bien complementario, pc_i y el nivel ingreso medio de los consumidores de este bien, I_i. Así, para estimar la demanda de este bien se emplean los siguientes modelos:

$$Q_i = \beta_0 + \beta_1 p_i + \beta_2 pc_i + \beta_3 ps_i + \beta_4 I_i + \varepsilon_i \tag{2.8}$$

$$\ln(Q_i) = \gamma_0 + \gamma_1 \ln(p_i) + \gamma_2 \ln(pc_i) + \gamma_3 \ln(ps_i) + \gamma_4 \ln(I_i) + \mu_i \tag{2.9}$$

$$Q_i = \beta_0 + \beta_1 p_i + \beta_2 pc_i + \beta_4 I_i + \varepsilon_i \tag{2.10}$$

$$Q_i = \beta_0 + \beta_1 p_i + \beta_2 ps_i + \beta_4 I_i + \varepsilon_i \tag{2.11}$$

Una vez estimados estos cuatro modelos se cuentan con su correspondiente R^2. Un "impulso natural" es escoger como el "mejor" modelo aquel que tenga el R^2 más grande. Pero tal procedimiento en este

caso es erróneo, pues al comparar el modelo 2.9 con los otros modelos nos damos cuenta que la variable dependiente es diferente en este modelo. Así, la SST del modelo 2.9 es distinta a la de los otros tres modelos, pues para el modelo 2.9 $SST = \sum_{i=1}^{n} \left(\ln\left(Q_i\right) - \ln\left(\bar{Q}\right)\right)^2$ mientras que para los otros modelos tenemos que $SST = \sum_{i=1}^{n} \left(Q_i - \bar{Q}\right)^2$.

Por lo tanto, si comparamos el R^2 del modelo 2.9 con el de los otros tres modelos, no estaríamos comparando la explicación de la misma variabilidad total de la variable dependiente. Podemos concluir que *los R^2 entre diferentes modelos son comparables si y solamente si la variable dependiente es la misma en los modelos a comparar.*

Pero, aun si comparamos modelos con la misma variable dependiente, como por ejemplo los modelos 2.8, 2.10 y 2.11, es necesario tener cuidado con esta comparación. Es relativamente fácil mostrar que el SSR, y por tanto el R^2, es una función creciente del número de regresores (k). Por tanto, a medida que consideramos modelos con más variables explicatorias, el R^2 será más grande.

Entonces, podríamos escoger cuál modelo explica mayor proporción de la variabilidad de la variable dependiente para los modelos 2.10 y 2.11 por medio del R^2, pues en estos dos modelos se tiene el mismo número de variables explicativas (k) y la variable explicada es la misma. Pero, en general, el R^2 no es un buen criterio para escoger entre los modelos 2.8, 2.10 y 2.11.

Para tener en cuenta la relación directa entre el SSR y el número de regresores (k), se ha diseñado el R^2 ajustado ($\bar{R}^2$). El $\bar{R}^2$ penaliza la inclusión de nuevas variables explicativas al modelo, esa penalización se da de la siguiente forma:

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

$\frac{n-1}{n-k}$ es el factor que penaliza el R^2 al incluir más variables explicativas, pues $\frac{n-1}{n-k}$ es creciente en k. En otras palabras, a medida que se incrementa el número de variables independientes en el modelo, $\frac{n-1}{n-k}$ aumenta al mismo tiempo que el R^2 aumenta y por tanto $\left(1 - R^2\right)$ disminuye. Así, el efecto de un aumento en el número de regresores no necesariamente implica un aumento en el $\bar{R}^2$. El $\bar{R}^2$ aumentará únicamente si el aumento en el R^2 es lo suficientemente grande para compensar la penalización que se hace por incluir más variables. Así un $\bar{R}^2$ grande será mejor que uno pequeño.

De esta manera, con el $\bar{R}^2$ se pueden comparar los modelos 2.8, 2.10 y 2.11. El único problema que se presenta con el $\bar{R}^2$, es que este carece de una interpretación así como la tiene el R^2.

Existen otros problemas, que se estudiarán más adelante, que pueden "inflar" el R^2. Así, aunque este estadístico presenta una interpretación muy clara e intuitiva, hay que tener cuidado antes de sacar conclusiones de este estadístico.

2.4. Pruebas conjuntas sobre los parámetros

Hasta aquí hemos discutido cómo comprobar hipótesis individuales en torno a cada uno de los parámetros de un modelo de regresión. Pero, ¿qué hacer con estos resultados individuales? Supongamos la siguiente situación; hemos estimado el modelo $y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$ con una muestra lo suficientemente grande. Además, hemos efectuado una serie de pruebas individuales sobre los parámetros estimados y se ha encontrado que:

- La hipótesis nula $\beta_2 = 0$ es rechazada a un nivel de significancia de α. Llamemos a este el resultado 1 (R_1)
- La hipótesis nula $\beta_3 = 0$ no se puede rechazar a un nivel de significancia de α. (R_2)
- La hipótesis nula $\beta_4 = 0$ no se puede rechazar a un nivel de significancia de α. (R_3)

Dados estos tres resultados, parece muy razonable tratar de unir estos tres resultados ($R_1 \cup R_2 \cup R_3$) y concluir que el modelo debería ser $y_i = \beta_1 + \beta_2 X_{2i} + \varepsilon_i$ con un nivel de significancia de α. Pero si reflexionamos un poco sobre las implicaciones de unir estos resultados, nos daremos cuenta rápidamente lo errado de unirlos. Noten que el primer resultado R_1 implica un error tipo I de tamaño α, mientras que los resultados R_2 y R_3 tienen asociados un error tipo II. Ahora, si unimos estos tres resultados, el nivel de significancia asociado a $(R_1 \cup R_2 \cup R_3)$ no será simplemente α.

Para evitar ser muy conservador al unir diferentes resultados individuales, se han diseñado pruebas que tengan en cuenta estos múltiples errores. Tales pruebas se conocen como pruebas conjuntas. En esta sección las discutiremos en detalle.

Supongamos que usted quiere determinar si todos los coeficientes asociados a pendientes son o no iguales a cero conjuntamente, es decir $H_0 : \beta_2 = \beta_3 = ... = \beta_k = 0$ versus la hipótesis alterna no H_o . Como lo discutimos anteriormente, este tipo de hipótesis no se puede probar con $k-1$ hipótesis individuales. Noten que esta hipótesis puede reescribirse de la forma $\mathbf{R_{(r)\times k}}\beta_{k\times 1} = \mathbf{C_{(r)\times 1}}$, donde $\beta^T = \left[\begin{array}{cccc} \beta_1 & \beta_2 & ... & \beta_k \end{array}\right]$, $\mathbf{C^T} = \left[\begin{array}{cccc} 0 & 0 & ... & 0 \end{array}\right]$ y:

$$R_{(r-1)\times k} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

Ahora, supongamos que quiere saber si dada una estimación de una función de producción Cobb-Douglas, ésta presenta rendimientos constantes de escala. Es decir, dada la siguiente ecuación estimada $\widehat{\ln(Y_i)} = \hat{\beta} + \hat{\alpha}_1 \ln(K_i) + \hat{\alpha}_2 \ln(L_i)$ se quiere comprobar si $\alpha_1 + \alpha_2$ es igual a uno o no. Esta hipótesis también se puede escribir de la forma $\mathbf{R_{(r)\times k}}\beta_{k\times 1} = \mathbf{C_{(r)\times 1}}$, donde $\beta^T = \left[\begin{array}{ccc} \alpha & \beta_2 & \beta_1 \end{array}\right]$, R = $\left[\begin{array}{ccc} 0 & 1 & 1 \end{array}\right]$ y $C = [1]$.

En general, cualquier hipótesis nula que pueda escribirse de la forma $R_{(r)\times k}\beta_{k\times 1} = C_{(r)\times 1}$ (versus la H_a: no H_0) se puede comprobar empleando el siguiente estadístico:

$$F_{calculado} = \frac{\left(\mathbf{C} - \mathbf{R}\hat{\beta}\right)^T \left(\mathbf{R}\left(\mathbf{X^T X}\right)^{-1}\mathbf{R^T}\right)^{-1} \left(\mathbf{C} - \mathbf{R}\hat{\beta}\right) \Big/ r}{\hat{\varepsilon}^T\hat{\varepsilon} \big/ (n-k)} \tag{2.12}$$

Este estadístico sigue una distribución F con r grados de libertad en el numerador y $n-k$ grados de libertad en el denominador. Por tanto se podrá rechazar la hipótesis nula, con un nivel de significancia α si $F_{\alpha,(r,(n-k))} < F_{Calculado}$.

Un caso especial de la hipótesis nula general $R_{(r)\times k}\beta_{k\times 1} = C_{(r)\times 1}$ (versus la H_a: no H_0) es $H_0 : \beta_2 = \beta_3 = ... = \beta_k = 0$ (todos los coeficientes asociados con pendientes son simultáneamente iguales a cero), es decir una prueba global de significancia. En este caso especial el estadístico $F_{calculado}$ se reduce a:

$$F_{Global} = \frac{\hat{\beta}^T\mathbf{X^T y} - n\bar{\mathbf{y}}^2 \Big/ (k-1)}{\hat{\varepsilon}^T\hat{\varepsilon} \big/ (n-k)} = \frac{MSR}{MSE} \tag{2.13}$$

O de manera equivalente:

$$F_{Global} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{2.14}$$

Este F_{Global} sigue una distribución F con $k-1$ grados de libertad en el numerador y $n-k$ grados de libertad en el denominador. Por tanto, se podrá rechazar la hipótesis nula a un nivel de significancia α si $F_{\alpha,((k-1),(n-k))} < F_{Global}$. Noten que este test se puede derivar rápidamente de la tabla ANOVA (Recuadro 2.3).

Recuadro 2.3 Tabla ANOVA con prueba de significancia global

Fuente de la Variación	**SS**	**Grados de libertad**	**MS**	**F-Global**
Regresión	SSR	$k-1$	$MSRE = \frac{SSRN}{k-1}$	$F_{Global} = \frac{MSRE}{MSR}$
Error	SSE	$n-k$	$MSR = s^2 = \frac{SSR}{n-k}$	
Total	SST	$n-1$		

Noten que la hipótesis nula de esta prueba de significancia global implica el modelo

$$y_i = \beta_1 + \varepsilon_i \tag{2.15}$$

mientras que la hipótesis alterna implica el modelo

$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \varepsilon_i \tag{2.16}$$

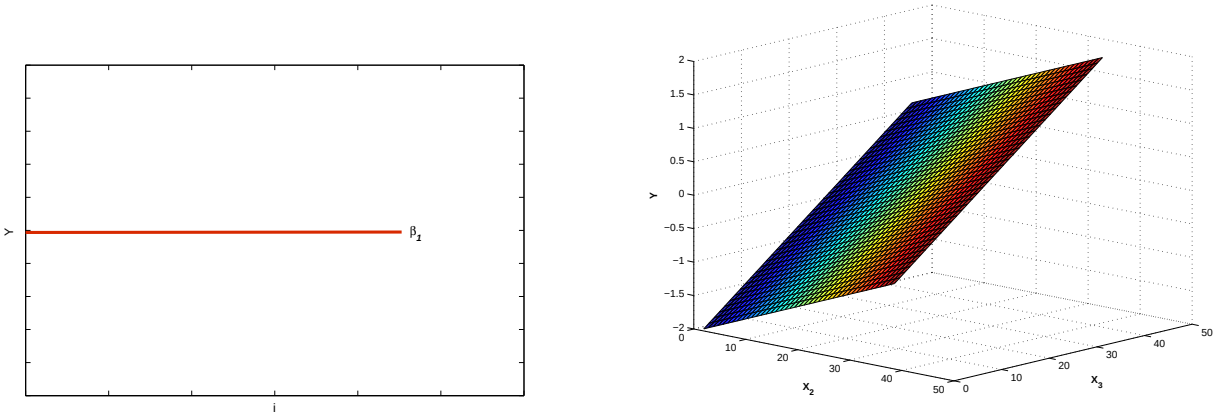
Es decir, si la hipótesis nula no es rechazada, esto implica que una mejor forma de explicar la variable dependiente es por medio de su media (una constante).[4] Por tanto, esta prueba es otra forma de estudiar la bondad de ajuste del modelo, por eso no debe ser sorprendente la estrecha relación entre el F_{Global} y el R_2.

$$F_{Global} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \tag{2.17}$$

[4]Es decir $Y_i = \mu + \varepsilon_i$, donde $\mu = \beta_1$ es la media.

Figura 2.1: Modelos 2.15 y 2.16

Modelo 2.15 **Modelo 2.16**



Otro caso especial es cuando consideramos una hipótesis nula en la cual un grupo de coeficientes[5] es conjuntamente igual a cero, es decir $H_o : \beta_p = \beta_{p+1} = \ldots = \beta_l = 0$, para $0 < p < l \leq k$ (versus la H_A: No H_0), esta prueba se conoce como una prueba de significancia conjunta. Es decir, la hipótesis nula es equivalente a $H_o : Y_i = \beta_1 + \beta_2 X_{2i} + \ldots + \beta_{p-1} X_{p-1i} + \beta_{l+1} X_{l+1i} + \ldots + \beta_k X_{ki} + \varepsilon_i$ (es decir un Modelo Restringido (R) del original) y la hipótesis alterna es $H_A : Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \varepsilon_i$ (el Modelo sin restricción (U). En este caso especial, el $F_{Calculado}$ de la expresión 2.12 es equivalente a:

$$F_c = \frac{(SSR_R - SSR_U)/r}{SSR_U/(n-k)}$$

donde $r = l - p$, SSR_R representa la suma de los cuadrados de los residuos estimados del modelo restringido y SSR_U representa la suma de los cuadrados de los residuos estimados del modelo sin restringir.

Así para probar la hipótesis nula que $H_o : \beta_p = \beta_{p+1} = \ldots = \beta_l = 0$, para $0 < p < l \leq k$ (versus la H_A: No H_0), podemos estimar por MCO el modelo restringido, implicado por la hipótesis nula, como el modelo sin restringir y encontrar su correspondiente SSE. A partir de estas cantidades se puede calcular el $F_{Calculado}$, el cual se compara con $F_{\alpha,(r,(n-k))}$. A este tipo de test se le conoce como una prueba de modelo restringido versus modelo no restringido.

Otro caso especial del test F para hipótesis de la forma $R\beta = C$, es conocido como la *prueba de Chow*[6] o prueba de cambio estructural. Test diseñado para detectar cambios en los parámetros del modelo estimado a lo largo de la muestra en estudio.

Suponga por ejemplo que desea estimar una función de consumo agregado de corte keynesiano, es decir el modelo a estimar es $C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$, donde C_t es el consumo agregado en el período t, Y_t es el ingreso agregado disponible en el período t y ε_t es un término aleatorio de error. Ahora suponga que usted espera un cambio estructural en el comportamiento del consumo agregado a partir del inicio de la década del 90. Es decir se espera un cambio tanto en el consumo autónomo, como en la propensión marginal a consumir. Así, nuestra hipótesis nula implicará un modelo en el cual no existe cambio estructural, es decir los parámetros del modelo son iguales antes y después de 1990

$$C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t, \forall t(R) \tag{2.18}$$

[5]Diferentes al término constante.

[6]Esta prueba recibe su nombre de su creador, Chow (1960)

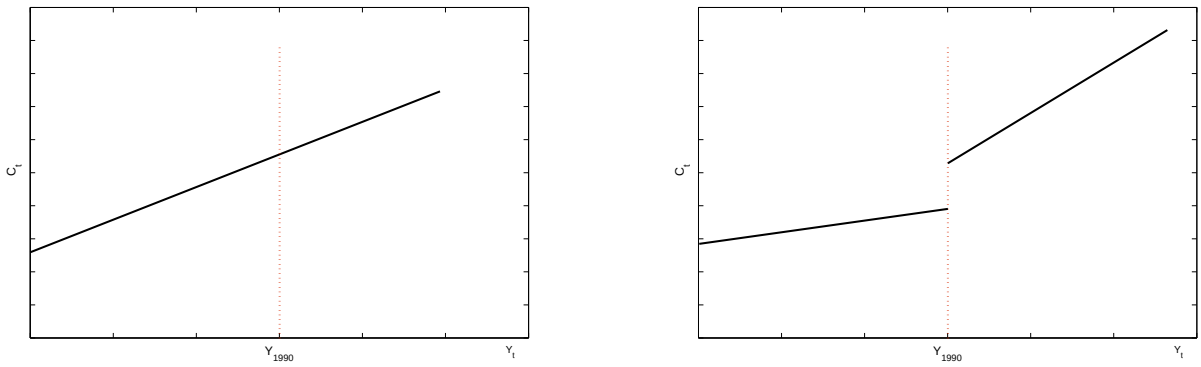
Mientras que bajo la hipótesis alterna, el modelo no posee restricción alguna, es decir los parámetros del modelo son diferentes en los dos períodos

$$\begin{aligned} C_t &= \beta_1 + \beta_2 Y_t + \varepsilon_t, \text{ para t}< 1990 \\ C_t &= \gamma_1 + \gamma_2 Y_t + \varepsilon_t, \text{ para t}\geq 1990. \end{aligned} \tag{2.19}$$

Esta hipótesis puede ser comprobada por medio de una prueba como la que hemos estado estudiando.

Figura 2.2: Modelos 2.18 y 2.25

Modelo 2.18 **Modelo 2.25**



En general, suponga que quiere probar la siguiente hipótesis nula $H_0 : Y_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i$ para $i = 1, 2, ...M + G$, es decir que no existe cambio estructural, versus la hipótesis alterna $Y_i = \gamma_1 + \gamma_2 X_{2i} + ... + \gamma_k X_{ki} + \varepsilon_i$ para $i = 1, 2, ...M$y

$$Y_i = \alpha_1 + \alpha_2 X_{2j} + ... + \alpha_k X_{kj} + \varepsilon_j$$

para $j = 1, 2, ...G$. Por tanto, el modelo bajo la hipótesis nula se puede considerar como un modelo restringido (R) pues este modelo implica $\gamma_1 = \alpha_1$,$\gamma_2 = \alpha_2$,...,$\gamma_k = \alpha_k$, el modelo con cambio estructural bajo la hipótesis alterna será pues un modelo sin restricciones. En este caso el estadístico $F_{Calculado}$ se convierte en:

$$F_c = \frac{(SSR_R - SSR_U)/k}{SSR_U/(n + M - 2k)}$$

Donde $SSR_U = SSR_1 + SSR_2$, SSR_1 corresponde al $SSRN$ de la regresión con los primeros M datos, y SSR_2 corresponde al SSR de la regresión con los restantes G datos. La hipótesis nula se rechazará si el F_C es mayor que el $F_{(k,n+M-2k)}$.

Recuadro 2.4 Prueba de Chow

El test de Chow consiste en los siguientes pasos. Es importante resaltar que el punto de quiebre debe ser preestablecido con anterioridad. Esta prueba no encuentra por si sólo el punto de quiebre o cambio estructural:

$H_0 : Y_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i$ (R) $i = 1, 2, ...M + G$
$H_A : Y_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \varepsilon_i$ (1) $i = 1, 2, ...M$
$H_A : Y_j = \alpha_1 + \alpha_2 X_{2j} + ... + \alpha_k X_{kj} + \varepsilon_j$ (2)(UR) $j = 1, 2, ...G$

1. Estime por MCO los tres modelos

2. Construya el siguiente estadístico: $F_c = \frac{(SSR_R - SSR_U)/k}{SSR_U/(n+M-2k)}$ con $SSRN_U = SSRN_1 + SSRN_2$

3. Rechace la hipótesis nula si $F_C < F_{(k,n+M-2k)}$

2.5. Prueba de Wald y su relación con la prueba F

En la sección anterior se discutió cómo para comprobar una hipótesis nula que involucre restricciones lineales de la forma $\mathbf{R}\hat{\beta} = \mathbf{C}$ podemos emplear el estadístico $F_{calculado}$ siguiente:

$$F_{calculado} = \frac{\left(\mathbf{C} - \mathbf{R}\hat{\beta}\right)^T \left(\mathbf{R}\left(\mathbf{X^T X}\right)^{-1}\mathbf{R^T}\right)^{-1} \left(\mathbf{C} - \mathbf{R}\hat{\beta}\right) \Big/ r}{\hat{\varepsilon}^T\hat{\varepsilon}/(n-k)} \tag{2.20}$$

Una alternativa para probar hipótesis que involucren restricciones de la forma $\mathbf{R}\hat{\beta} = \mathbf{C}$ es emplear el estadístico de Wald ($W_{calculado}$), el cual está relacionado con el $F_{calculado}$ descrito con anterioridad. Éste estadístico se puede calcular con la siguiente expresión:

$$W_{calculado} = \left(\mathbf{C} - \mathbf{R}\hat{\beta}\right)^T \left(\mathbf{R}\left(S^2\mathbf{X^T X}\right)^{-1}\mathbf{R^T}\right)^{-1} \left(\mathbf{C} - \mathbf{R}\hat{\beta}\right) \tag{2.21}$$

El estadístico de Wald calculado sigue una distribución χ^2_r donde r representa el número de restricciones lineales en la hipótesis nula. Por tanto, se podrá rechazar $H_0 : \mathbf{R}\beta = \mathbf{C}$ en favor de la hipótesis alterna (no H_0) si $W_{calculado} > \chi^2_r$.

Es muy fácil encontrar la relación entre el estadístico de Wald y el estadístico F. Por ejemplo, manipulando algebraicamente (2.21) tendremos que:

$$W_{calculado} = \left(\mathbf{C} - \mathbf{R}\hat{\beta}\right)^T \left(\mathbf{R}\left(\mathbf{X^T X}\right)^{-1}\mathbf{R^T}\right)^{-1} \left(\mathbf{C} - \mathbf{R}\hat{\beta}\right) \left(S^2\right)^{-1} \tag{2.22}$$

$$W_{calculado} = \frac{\left(\mathbf{C} - \mathbf{R}\hat{\beta}\right)^T \left(\mathbf{R}\left(\mathbf{X^T X}\right)^{-1}\mathbf{R^T}\right)^{-1} \left(\mathbf{C} - \mathbf{R}\hat{\beta}\right)}{S^2} \tag{2.23}$$

y reemplazando S^2 tenemos que:

$$W_{calculado} = \frac{\left(\mathbf{C} - \mathbf{R}\hat{\beta}\right)^T \left(\mathbf{R}\left(\mathbf{X^T X}\right)^{-1}\mathbf{R^T}\right)^{-1} \left(\mathbf{C} - \mathbf{R}\hat{\beta}\right)}{\hat{\varepsilon}^T\hat{\varepsilon}/(n-k)} \tag{2.24}$$

Al comparar la anterior expresión con (2.20) se puede concluir que $W_{calculado} = r \cdot F_{calculado}$. De hecho, para comprobar cualquier hipótesis que involucre restricciones lineales de la forma $\mathbf{R}\hat{\beta} = \mathbf{C}$ se podrá emplear una prueba F o una prueba Wald. Sin importar el estadístico que se emplee, la conclusión será la misma. En el caso de EasyReg, es importante resaltar que este software calcula únicamente el estadístico Wald.

2.6. Restricciones lineales sobre los parámetros con EasyReg

EasyReg nos permite probar fácilmente cualquier tipo de restricción lineal de los parámetros de la forma $R\beta = C$ con un estadístico basado en una versión abreviada de la razón de verosimilitud (Likelihood Ratio Test) que sigue una distribución Chi-cuadrado con grados de libertad igual al número de restricciones lineales. Este tipo de prueba es conocida como test de Wald.

En esta sección estudiaremos la productividad de los factores (PTF) para Colombia, para ello utilizaremos la ecuación neoclásica básica del crecimiento:

$$Y = cL^{\alpha_1}K^{\beta_1}$$

Linealizando tenemos:

$$Ln(Y) = Ln(c) + \alpha_1 Ln(L) + \beta_1 Ln(K)$$

donde,

$$Ln(c) = \beta_0$$

Finalmente obtenemos el siguiente modelo a estimar:

$$Ln(Y_t) = \beta_0 + \alpha_1 Ln(L_t) + \beta_1 Ln(K_t) + \varepsilon_t \tag{2.25}$$

donde Y_t es el producto, L_t es el empleo y K_t el stock de Capital. Para esto contamos con datos trimestrales desestacionalizados desde el primer trimestre de 1980 hasta el primer trimestre de 1996. Los datos corresponden a Alonso y Cárdenas (1997).

Para empezar con nuestro ejercicio cargue los datos que se encuentran en el archivo cobbdouglas.xls, luego transforme las series Y_t, L_t, K_t en logaritmos. Posteriormente estime el modelo 2.25. Obtendrá los resultados del cuadro 2.1.

Cuadro 2.1: Estimación del modelo 2.25

	Variable dependiente $LN(Y_t)$ Estadísticos t entre paréntesis
	Ecuación 2.25 1980:1 - 1996:1 MCO
Constante	-1.2153181 (-2.560)**
$LN(K_t)$	0.6514469 (28.781)***
$LN(L_t)$	0.3940024 (7.140)***
R^2	0.9965
$R^2 ajustado$	0.9964
F-Global	5603.06***
No. de obs.	42

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos Cuadrados Ordinarios

En el siguiente cuadro se presenta la tabla Anova del modelo estimado.

Cuadro 2.2: Tabla Anova del modelo 2.25 estimado

Fuente de la Variación	**SS**	**Grados de libertad**	**MS**	**F-Global**
Regresión	14.98616413	2	7.493082065	5603.059478
Error	0.05215547	39	0.00133732	
Total	15.0383196	41		

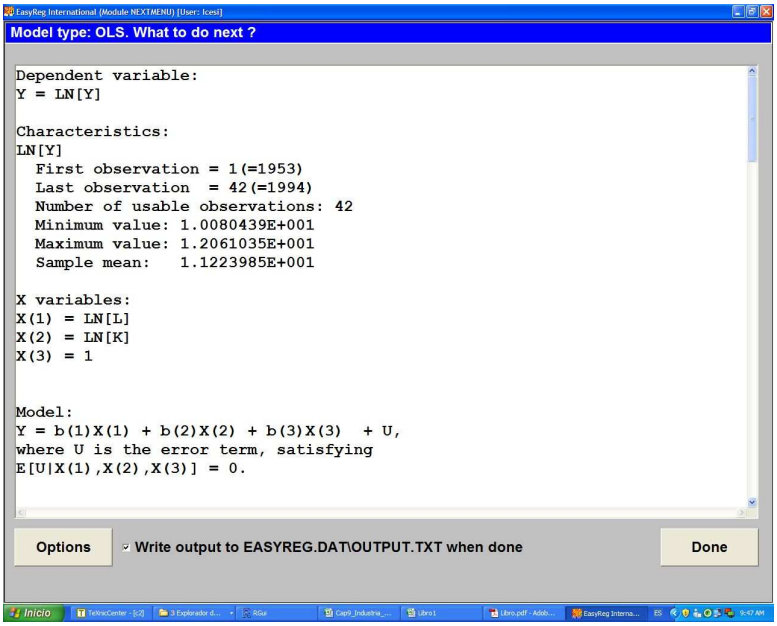
El R^2 de esta regresión indica que el modelo estimado explica en un 99.65 % la variación del logaritmo natural del producto. Además, es posible rechazar la hipótesis nula que los coeficientes asociados a la constante, logarítmo natural del empleo y al logaritmo natural del stock de capital son individualmente iguales a cero, con un nivel de significancia del 5 % para la constante y del 1 % en el caso de las pendientes. Finalmente en el cuadro 2.1 podemos observar que el estadístico F-Global corresponde a 5603.06, con el cual es posible rechazar la hipótesis nula que las pendientes del modelo son conjuntamnete iguales a cero, con un nivel de significancia del 1 %.

Ahora, consideremos las siguientes pruebas de hipótesis:

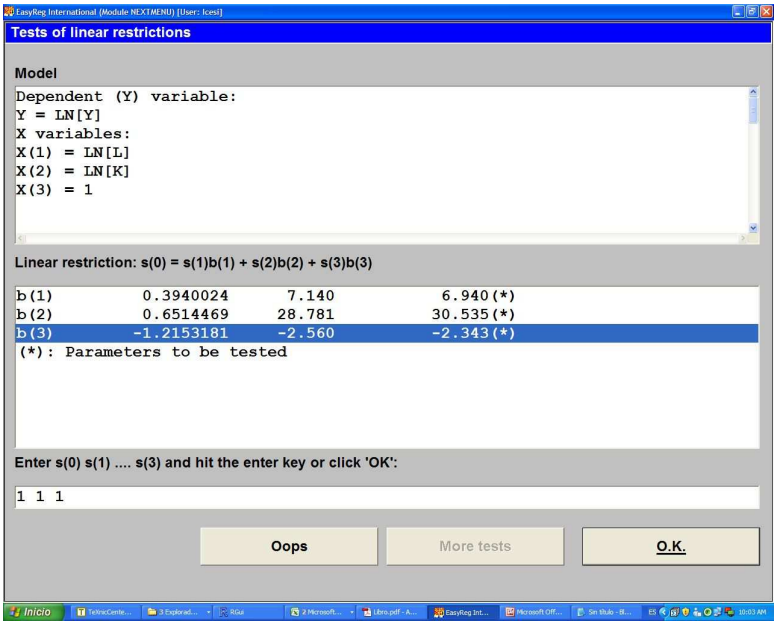
1. $H_0 : \beta_0 = \alpha_1 = \beta_1 = 0$ (esta prueba de hipótesis no tiene sentido económico, sólo se hace para ilustrar la técnica)
2. $H_0 : \alpha_1 + \beta_1 = 1$

Mediante la prueba 1) Determinaremos si todos los parámetros del modelo (incluyendo el intercepto) son conjuntamente significativos. Finalmente, con la prueba 2) comprobaremos si en Colombia durante el periodo estudiado, existen rendimientos constantes de los factores.

La prueba de hipótesis 1) $H_0 : \beta_0 = \alpha_1 = \beta_1 = 0$ la podemos llevar a cabo al finalizar la estimación haciendo clic en "Options/Wald test of linear parameter restrictions".



La siguiente ventana nos permite escoger los coeficientes involucrados en la prueba a realizar. Para ello haga doble clic en las líneas correspondientes a b(1), b(2) y b(3); estos coeficientes corresponden a los estimadores MCO para β_1, α_1 y β_0 respectivamente. Una vez seleccionados estos coeficientes, haga clic en el botón "Test joint significance" (en general esta opción prueba la hipótesis nula que los coeficientes seleccionados son conjuntamente iguales a cero).



El resultado de este cálculo aparecerá en la ventana siguiente.

Recuadro 2.5 Resultados de la prueba de Wald para $H_0 : \beta_0 = \alpha_1 = \beta_1 = 0$

Wald test:

b(1)	0.3940024	7.140	6.940(*)
b(2)	0.6514469	28.781	30.535(*)
b(3)	-1.2153181	-2.560	-2.343(*)

(*): Parameters to be tested

Null hypothesis:
b(1) = b(2) = b(3) = 0

Wald test: 3967679.28
Asymptotic null distribution: Chi-square(3)
p-value = 0.00000

Significance levels:	10 %	5 %
Critical values:	6.25	7.81
Conclusions:	reject	reject

N.B.: The latter results are based on the Chi-square distribution.

Test result on the basis of the heteroskedasticity consistent variance matrix:
Wald test: 4317727.24
Asymptotic null distribution: Chi-square(3)
p-value = 0.00000

Significance levels:	10 %	5 %
Critical values:	6.25	7.81
Conclusions:	reject	reject

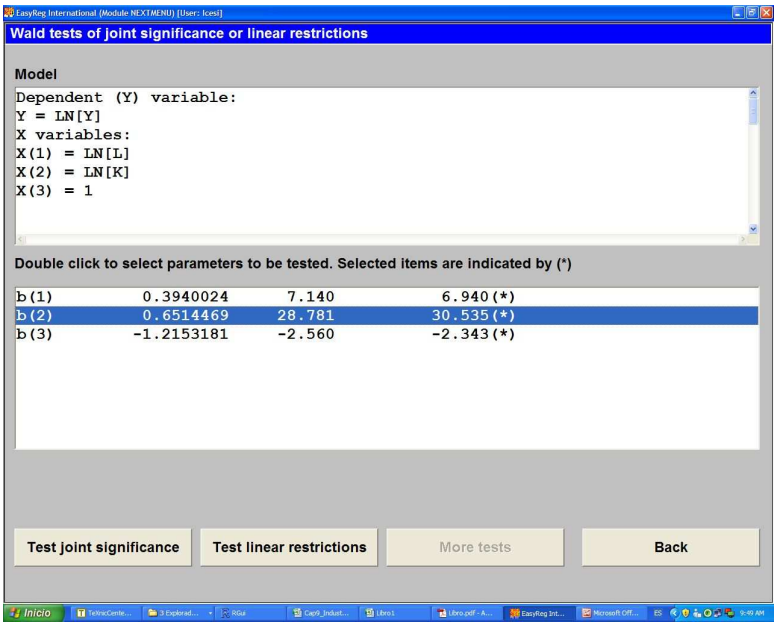
El estadístico Wald para este test es 3967679.28, este estadístico sigue una distribución Chi-cuadrado con tres grados de libertad. Dado que el correspondiente valor p es menor que 0.01, entonces es posible rechazar la hipótesis nula que $\alpha_1 = 0$, $\beta_1 = 0$ y $\beta_0 = 0$ son al mismo tiempo iguales a cero con un nivel de significancia del uno por ciento.

Ahora probemos la hipótesis 2). Para esto, haga clic en el botón "More tests". Seleccione los dos coeficientes haciendo doble clic sobre ellos y haga clic en el botón "Test linear restrictions". Note que la $H_0 : \alpha_1 + \beta_1 = 1$ se puede escribir de la forma $R\beta = C$, donde:

$$\mathbf{R} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \beta = \begin{bmatrix} \alpha_1 \\ \beta_1 \end{bmatrix} \text{ y } \mathbf{C} = [\mathbf{1}].$$

Además, recuerde que los estimadores para α_1 y β_1 calculados por EasyReg son denominados como b(2) y b(1), respectivamente.

Volviendo a EasyReg, escoja la opción "Wald test of linear parameters restrictions", escoja los coeficientes involucrados y haga clic en botón "Test linear restrictions". El software requiere que escribamos cada una de las restricciones como una línea (es decir fila por fila de la matriz $R \cdot \beta$, en la forma $s(1) \cdot b(1) + s(2) \cdot b(2) + s(3) \cdot b(3) + ... = s(0)$, así para la restricción $H_0 : \alpha_1 + \beta_1 = 1$ tenemos que $c(0) = 1$, $c(1) = 1$ y $c(2) = 1$, Ahora escriba "1 1 1" en la ventana "Enter s(0) s(1) s(2) and..." selecciones "'Ok' y luego "No more restrictions"



El resultado aparecerá en la siguiente ventana. Haga clic en el botón "Back". Los resultados de este cálculo se pueden ver en el recuadro 2.6. Note que EasyReg escribe las correspondientes matrices **R** y **C**. En este caso el estadístico Wald es 1.71 que está distribuido Chi-cuadrado con 1 grado de libertad. Los valores críticos con niveles de significancia 10 % y 5 % son 2.71 y 3.84, respectivamente. Entonces no es posible rechazar la hipótesis nula ni siquiera al 10 % de significancia. En conclusión no hay suficiente evidencia para rechazar la hipótesis nula. Asi es posible afirmar que para los datos empleados y para el periodo estudiado el modelo 2.25 muestra rendimientos constantes para los factores trabajo y capital.

Recuadro 2.6 Resultados del test de Wald para $H_0 : \alpha_1 + \beta_1 = 1$

Wald test:

b(1)	0.3940024	7.140	6.940(*)
b(2)	0.6514469	28.781	30.535(*)
b(3)	-1.2153181	-2.560	-2.343

(*): Parameters to be tested

Null hypothesis:
1.b(1)+1.b(2) = 1.

Null hypothesis in matrix form: Rb = c, where
R =
1. 1. 0.
and c =
1.
Wald test on the basis of the standard variance matrix:
Wald test: 1.71
Asymptotic null distribution: Chi-square(1)
p-value = 0.19037

Significance levels:	10 %	5 %
Critical values:	2.71	3.84
Conclusions:	accept	accept

Wald test on the basis of White's heteroskedasticity consistent
variance matrix:
Wald test statistic: 1.41
Asymptotic null distribution: Chi-square(1)
p-value = 0.00309

Significance levels:	10 %	5 %
Critical values:	2.71	3.84
Conclusions:	accept	accept

2.7. Ejercicios

Continuando con el ejercicio del Capítulo 1, responda:

1. El modelo: $I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 LDies_t + \alpha_5 LEI_t + \alpha_6 V_t + \varepsilon_t$ generó mucha discusión en la pequeña república y en diversos foros se sugirieron diferentes modelos:

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \varepsilon_t \tag{2.26}$$

$$I_t = \alpha_1 + \alpha_2 LDies_t + \alpha_3 LEI_t + \varepsilon_t \tag{2.27}$$

$$I_t = \alpha_1 + \alpha_2 V_t + \varepsilon_t \tag{2.28}$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 LEI_t + \varepsilon_t \tag{2.29}$$

$$I_t = \alpha_1 + \alpha_2 CD_t + \alpha_3 LDies_t + \varepsilon_t \tag{2.30}$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 LDies_t + \alpha_4 LEI_t + \alpha_5 V_t + \varepsilon_t \tag{2.31}$$

$$I_t = \alpha_1 + \alpha_2 CE_t + \alpha_3 CD_t + \alpha_4 V_t + \varepsilon_t \tag{2.32}$$

   Determine cuál es el mejor modelo para explicar los ingresos del sector ferroviario en esta pequeña república (muestre todos los cálculos necesarios para tomar esta decisión y asegúrese que su respuesta está bien argumentada).

2. A partir del modelo que seleccionó en la pregunta anterior:

   a) Construya la tabla ANOVA.

   b) Analice la significancia individual de los coeficientes

   c) Analice la significancia conjunta de los coeficientes

   d) Determine cuál es el factor que más afecta el ingreso del sector ferroviario en esta nación

   e) De acuerdo con su resultado, ¿qué puede sugerir al gobierno de esta nación para mejorar los ingresos del sector?

2.8. Apéndice

Apéndice 2.1 Demostración de la ecuación 2.7

La expresión 2.7 implica que en presencia de un modelo lineal con intercepto, podremos descomponer la variación total de la variable dependiente (SST) en la parte explicada por el modelo de regresión (SSR) y la parte no explicada por el modelo (SSE). Formalmente, la proposición a probar es que si una de las variables explicativas es una constante (corresponde al intercepto), entonces:

$$SST = SSR + SSE$$

Para demostrar esta proposición, podemos partir del hecho que:

$$\hat{\varepsilon}^T\hat{\varepsilon} = \left(y - \mathbf{X}\hat{\beta}\right)^T\left(y - \mathbf{X}\hat{\beta}\right)$$

$$\hat{\varepsilon}^T\hat{\varepsilon} = y^Ty - 2\hat{\beta}^T\mathbf{X}^Ty + \hat{\beta}^T\mathbf{X}^T\mathbf{X}\hat{\beta}$$

$$\hat{\varepsilon}^T\hat{\varepsilon} = y^Ty - \hat{\beta}^T\mathbf{X}^T\mathbf{X}\hat{\beta}$$

Remplazando, obtenemos

$$y^Ty = \hat{\varepsilon}^T\hat{\varepsilon} + \hat{y}^T\hat{y}$$

Restando a ambos lados $n\bar{y}^2$, se obtiene:

$$y^Ty - n\bar{y}^2 = \hat{\varepsilon}^T\hat{\varepsilon} - n\bar{y}^2 + \hat{y}^T\hat{y}$$

En otras palabras, tenemos que:

$$SST = SSE + \hat{y}^T\hat{y} - n\bar{y}^2$$

Ahora, será necesario demostrar que $SSR = \hat{y}^T\hat{y} - n\bar{y}^2$ siempre que el modelo lineal incluya un intercepto. Noten que:

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}\hat{y}_i^2 - 2\sum_{i=1}^{n}\hat{y}_i\bar{y} + \sum_{i=1}^{n}\bar{y}^2$$

$$SSR = \hat{y}^T\hat{y} - 2\bar{y}\sum_{i=1}^{n}\hat{y}_i + n\bar{y}^2$$

Remplazando el valor estimado de la variable dependiente, se tiene que:

$$SSR = \hat{y}^T\hat{y} - 2\bar{y}\sum_{i=1}^{n}(y_i - \hat{\varepsilon}_i) + n\bar{y}^2$$

$$SSR = \hat{y}^T\hat{y} - 2\bar{y}\left[\sum_{i=1}^{n}y_i - \sum_{i=1}^{n}\hat{\varepsilon}_i\right] + n\bar{y}^2$$

$$SSR = \hat{y}^T\hat{y} - 2\bar{y}n\bar{y} + 2\bar{y}\sum_{i=1}^{n}\hat{\varepsilon}_i + n\bar{y}^2$$

Recordemos que si uno de los regresores es una constante, entonces ya se demostró que siempre se cumplirá que $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$. Por eso,

$$SSR = \hat{y}^T \hat{y} - 2n\bar{y}^2 + n\bar{y}^2$$

$$SSR = \hat{y}^T \hat{y} - n\bar{y}^2$$

Es decir, se tiene que:

$$SST = SSE + SSR$$

Es importante anotar que si no se garantiza que $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$, entonces no necesariamente se puede garantizar que la suma del SSR y el SSE sean iguales al SST. Y esto solo se puede asegurar cuando el modelo estimado tiene un intercepto.

Apéndice 2.2 Prueba de no normalidad Jarque-Bera/Salmon-Kiefer para los errores estimados

Como se discutió anteriormente, en caso que la muestra no sea lo suficientemente grande, entonces para hacer inferencia será necesario suponer que el término aleatorio de error sigue una distribución normal. A continuación se discute la prueba de no normalidad para los residuos estimados de Jarque-Bera/Salmon-Kiefer que es calculada automáticamente por EasyReg. La idea detrás de esta prueba es muy sencilla. En general, si una muestra aleatoria proviene de una variable aleatoria que sigue una distribución normal, entonces la muestra debe tener las características de una distribución normal. Es decir, la muestra debe tener una distribución simétrica (coeficiente de asimetría (A) igual a cero) y además la curtosis (K) debe ser igual a 3.

La prueba de Jarque-Bera/Salmon-Kiefer tiene como hipótesis nula que los residuos estimados provienen de una población de errores que sigue una distribución normal ($H_0 : \hat{\varepsilon} \sim N$) versus la alterna de no H_0. El estadístico Jarque-Bera/Salmon-Kiefer (JB) que permite comprobar dicha hipótesis corresponde a:

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

Este estadístico debe compararse χ^2_2 y en caso de que $JB > \chi^2_2$, se podrá rechazar la hipótesis nula. EasyReg siempre reporta este estadístico, el valor crítico de la distribución Chi-cuadrado (para los niveles de significancia del 10 % y del 5 %) y el respectivo valor-p. Por ejemplo, en el recuadro 4 del capítulo donde se presenta una salida de EasyReg se puede encontrar que el estadístico JB es de 3.873718 con un valor-p de 0.14416. Por lo tanto no es posible rechazar la hipótesis nula de que los residuos estimados provienen de una población que sigue una distribución normal.

CAPÍTULO 3

Variables dummy

Objetivos del capítulo

Al terminar la lectura de este capítulo el lector estará en capacidad de:

- Crear variables dummy con EasyReg.
- Estimar modelos econométricos con variables dummy que permitan comprobar diferentes hipótesis y casos.

3.1. Introducción

En los capítulos anteriores hemos discutido el uso de variables que toman valores en un dominio continuo en los modelos de regresión lineal. Sin embargo, en algunas ocasiones será necesario emplear variables que toman únicamente dos valores distintos. Por ejemplo, existirán algunas ocasiones en que para la estimación de un modelo sería relevante considerar si un individuo es mujer o no o si es de un país que pertenece a determinado grupo económico o no.

En este capítulo vamos a discutir las variables dummy, también conocidas como variables ficticias o variables dicotómicas.

Ejemplo 3.1 Variables dummy

Suponga que un equipo de fútbol presenta dos volúmenes de ventas de entradas a los partidos(Y_i) bastante distintos:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \quad X_i = \begin{cases} 1 & \text{Si juega una estrella en el equipo contrario} \\ 0 & \text{En caso contrario (o.w)} \end{cases}$$

Por lo tanto, si juega una súper estrella en el equipo contrario tenemos que el valor esperado de las ventas de boletos corresponde a: $E[Y_i] = E[\beta_1 + \beta_2 X_i + \varepsilon_i]$ y aplicando las propiedades del valor esperado obtenemos: $E[Y_i] = \beta_1 + E[\beta_2 X_i]$ y finalmente $E[Y_i] = \beta_1 + \beta_2 E[X_i] = \beta_1 + \beta_2$.

Mientras que si no juega una súper estrella las ventas de boletos están dadas por: $E[Y_i] = E[\beta_1 + \beta_2 X_i + \varepsilon_i]$. Aplicando nuevamente las propiedades del valor esperado tenemos que $E[Y_i] = \beta_1 + E[\beta_2 X_i]$ y finalmente al ser X_i igual a 0, obtenemos: $E[Y_i] = \beta_1 + \beta_2 E[X_i] = \beta_1$.

Podemos concluir entonces que las ventas de boletos están explicadas por:

$$Y_i = \begin{cases} \beta_1 + \beta_2 + \varepsilon_i & \text{Si juega una estrella en el equipo contrario} \\ \beta_1 + \varepsilon_i & \text{En caso contrario (o.w)} \end{cases}$$

donde β_2 representa la diferencia que presentan las ventas de entradas a los partidos cuando en el equipo contrario juega una súper estrella.

3.2. Usos de las variables dummy

En la práctica los modelos econométricos que incluyen variables dummy también contienen en su especificación variables continuas simultáneamente. Para comprender mejor el uso de las variables ficticias emplearemos un ejemplo más general, que consiste en la función de consumo keynesiana que se comporta de manera diferente en tiempos de contracción económica que en años de expansión. En este ejemplo podemos plantear los cuatro casos posibles en que el consumo agregado está determinado por el ingreso disponible.

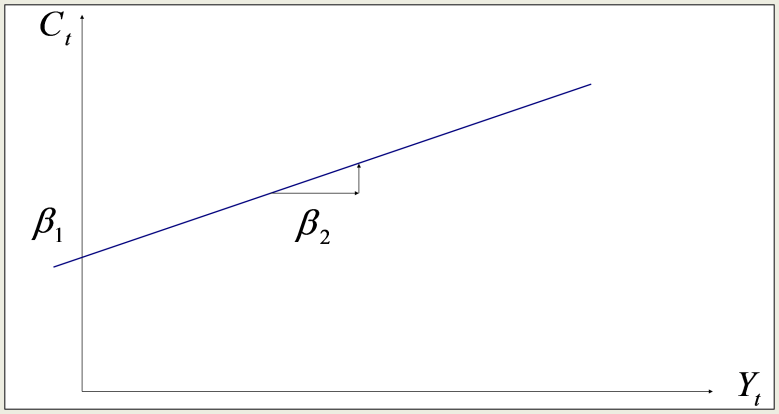
Ejemplo 3.2 Función de consumo keynesiana

Caso I

En este caso se asume que la función de consumo es igual tanto en tiempos de contracción como en tiempos de expansión. Como podemos apreciar en la gráfica tanto la pendiente como el intercepto de la función, se mantienen inalterados. Por lo tanto, la función de consumo para ambas circunstancias está dada por:

$$C_t = \beta_1 + \beta_2 Y_t + \varepsilon_t$$

Gráficamente tenemos:



En este caso β_2 representa la propensión marginal a consumir y β_1 el consumo autónomo.

Caso II

En el segundo caso, tenemos que el consumo es menor en tiempo de contracción que en tiempos de expansión (ceteris paribus). Esto se puede representar con una reducción del consumo autónomo (intercepto de la función) pero manteniéndose inalterada la propensión marginal al consumo (pendiente). Este efecto lo recoge el siguiente modelo:

$$C_t = \beta_1 + \beta_2 Y_t + \alpha D_t + \varepsilon_t \quad \text{donde} \quad \mathrm{D_t} = \begin{cases} 1 & \text{periodos de contracción} \\ 0 & o.w. \end{cases} \tag{3.1}$$

Como podemos apreciar, la ecuación (3.1) sugiere que en periodos de contracción el valor que toma la variable dummy es de 1. Por lo tanto, en tiempo de contracción el valor esperado del consumo está dado por:

$$E\left[C_t\right] = (\beta_1 + \alpha) + \beta_2 Y_t$$

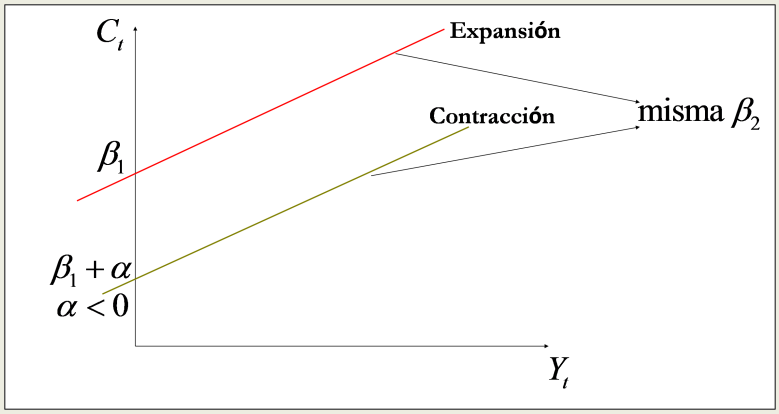
En tiempos de expansión, la variable dummy toma el valor de 0 y por lo tanto el valor esperado del consumo se reduce a:

$$E\left[C_t\right] = \beta_1 + \beta_2 Y_t$$

Finalmente, podemos representar el valor esperado del consumo capturando el cambio en el intercepto y con pendiente constante de la siguiente manera:

$$E\left[C_t\right] = \begin{cases} (\beta_1 + \alpha) + \beta_2 Y_t & \text{periodos de contracción} \\ \beta_1 + \beta_2 Y_t & o.w. \end{cases}$$

Gráficamente tenemos:



En este caso, α representa la diferencia del consumo autónomo entre el periodo de contracción y de expansión. Además, se espera que $\alpha > 0$

Caso III

En este caso, el consumo es nuevamente menor en tiempo de contracción que en tiempos de expansión, pero a diferencia del Caso II en el que solo cambiaba el intercepto de la función, en este caso se presenta un cambio de pendiente (los individuos consumen menos por cada unidad monetaria adicional de ingreso disponible). Este hecho lo podemos representar con el siguiente modelo:

$$C_t = \beta_1 + \beta_2 Y_t + \gamma\left(D_t Y_t\right) + \varepsilon_t \quad \text{donde} \quad \mathrm{D_t} = \begin{cases} 1 & \text{periodos de contracción} \\ 0 & o.w. \end{cases} \tag{3.2}$$

La ecuación (3.3) nos muestra que en tiempo de contracción el valor que toma la variable dummy es de 1, produciéndose un cambio en pendiente. Es decir, tenemos que en tiempos de contracción el valor esperado del consumo se especifica de la siguiente forma:

$$E\left[C_t\right] = \beta_1 + \left(\beta_2 + \gamma\right) Y_t$$

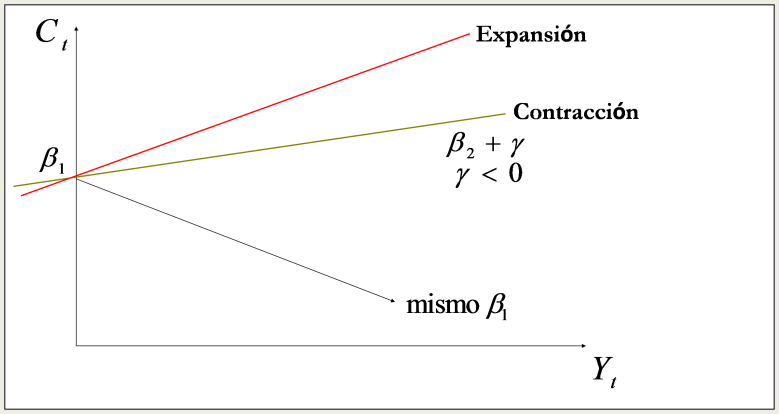
Al igual que en el caso anterior, en tiempos de expansión la variable dummy toma el valor de 0 y por lo tanto el valor esperado del consumo viene dado por:

$$E\left[C_t\right] = \beta_1 + \beta_2 Y_t$$

Finalmente podemos representar el valor esperado del consumo capturando el cambio en la pendiente de la siguiente forma:

$$E\left[C_t\right] = \begin{cases} \beta_1 + \left(\beta_2 + \gamma\right) Y_t & \text{periodos de contracción} \\ \beta_1 + \beta_2 Y_t & \text{o.w.} \end{cases}$$

Gráficamente tenemos:



En este caso γ corresponde a la diferencia entre la propensión marginal a consumir en periodos de contracción y expansión. En este caso se espera que $\gamma > 0$

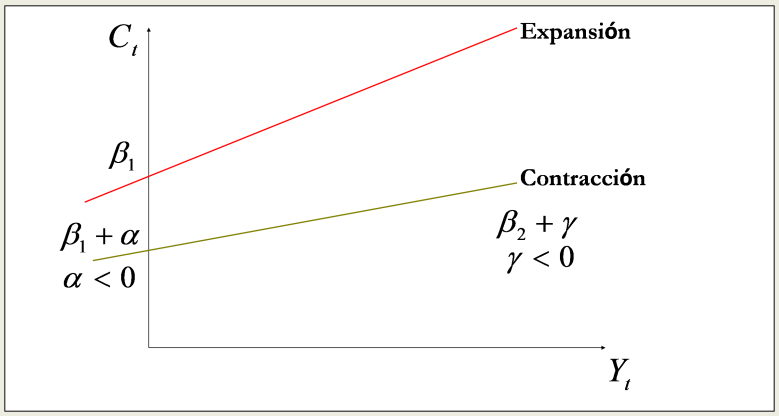
Caso IV

En este último caso, el consumo se reduce en tiempos de contracción por las dos posibles vías, es decir, por una reducción tanto en el intercepto como en la pendiente. Este efecto es capturado por un modelo con la siguiente especificación:

$$C_t = \beta_1 + \beta_2 Y_t + \alpha D_t + \gamma (D_t Y_t) + \varepsilon_t \quad \text{donde} \quad \mathrm{D_t} = \begin{cases} 1 & \text{periodos de contracción} \\ 0 & o.w. \end{cases} \tag{3.3}$$

Finalmente, podemos representar el valor esperado del consumo capturando el cambio en la pendiente y en el intercepto de la siguiente forma:

$$E[C_t] = \begin{cases} (\beta_1 + \alpha) + (\beta_2 + \gamma) Y_t & \text{periodos de contracción} \\ \beta_1 + \beta_2 Y_t & \text{o.w.} \end{cases}$$



En este caso α y γ representan respectivamente la diferencia del consumo autónomo y de la propensión marginal a consumir en periodos de contracción y de expansión. Se espera que α y γ sean negativos.

3.3. Relación entre la economía mundial y la colombiana

Con este ejercicio ilustraremos cómo generar variables dummy con EasyReg. Nos basaremos en la Ley Thirlwall formulada en 1979, la cual implica que para una economía abierta, la demanda impone restricciones al crecimiento dadas las diferentes dinámicas que ésta presenta en cada uno de los países y que ocasiona que crezcan a tasas diferentes.[1]

El modelo para probar la Ley de Thirlwall implica un modelo lineal del logaritmo del PIB (Y_t) del país bajo estudio y en función del logaritmo del PIB del resto del mundo (f_t), es decir:

$$Y_t = \beta_0 + \beta_1 f_t + \varepsilon_t \tag{3.4}$$

donde β_0 representa el intercepto, β_1 es la razón entre las elasticidades ingreso de las exportaciones e importaciones y ε_t es un término de error aleatorio.

Es más, Thirlwall demuestra que si β_1 fuera igual a 1, esto significaría que las tasas de crecimiento de largo plazo de la economía en estudio y las del resto de la economía son similares. Por otra parte, se tiene que si β_1 es mayor que 1, entonces el desempeño de la economía en estudio está por encima de la del resto del mundo, este es el caso deseable. El resultado es contrario cuando β_1 es menor que 1. De igual manera, se tiene que si β_1 es mayor que 1, entonces la elasticidad ingreso de las exportaciones es superior a la de las importaciones; por tanto, a medida que se incrementa el ingreso de las familias, una mayor proporción del incremento del ingreso se destina a las importaciones. Desde otro punto de vista, esto implica que si la tasa de crecimiento del resto del mundo aumenta, entonces se puede elevar el crecimiento interno de la economía nacional en una mayor cuantía que el incremento que se dio en el resto del mundo. Por otra parte, si $0<\beta_1<1$, entonces ante incrementos en el nivel de crecimiento de la economía mundial, la economía doméstica crecerá en menor cuantía a este aumento. Esta es otra razón por la cual es preferible que β_1 se aleje de 1.

Para estimar la ecuación (3.4) es necesario seleccionar una variable que refleje el comportamiento de la economía mundial, en nuestro caso seleccionamos el logaritmo del PIB de Estados Unidos (f_t) como variable "proxy".[2] Los datos de los logaritmos del PIB colombiano y el de Estados Unidos se encuentran en el archivo edummy.xls.

Si graficamos los datos por medio de un gráfico de dispersión podemos ver que parece haber un cambio en la relación entre el PIB de Colombia con respecto al de Estados Unidos al comenzar la década de los noventa (figura 3.1). Sin embargo, el análisis gráfico no es suficiente para determinar si en efecto existe un cambio en el intercepto, en la pendiente o en ambos.

Para verificar la hipótesis de que la relación ha cambiado después de 1990 (periodo de la llamada apertura económica) se considera el siguiente modelo

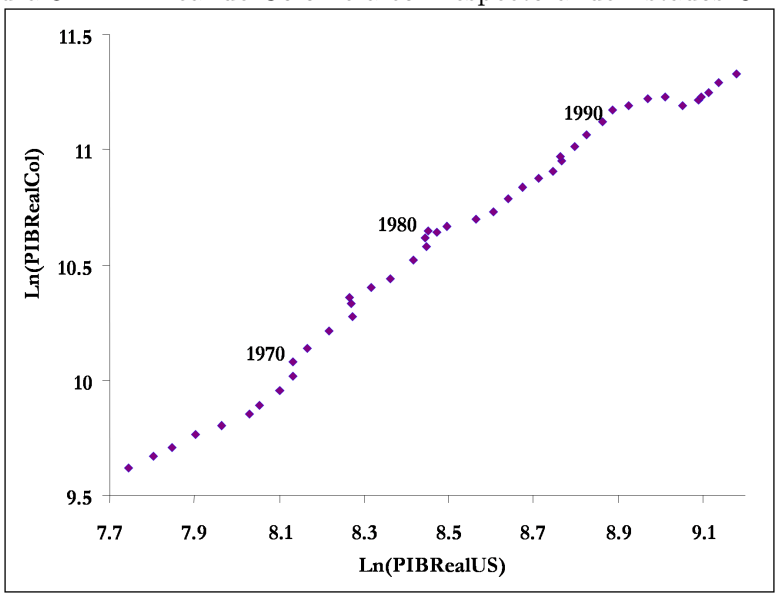
$$Y_t = \beta_0 + \beta_1 f_t + \alpha_2 D_t + \beta_2 (D_t f_t) + e_t \quad \text{donde} \quad \mathrm{D_t} = \begin{cases} 1 & si \;\; t \geq 1991 \\ 0 & o.w. \end{cases}$$

Para estimar este modelo, el primer paso es leer los datos en EasyReg.

[1] Para mayor información ver Thirlwall (1979).

[2] Por variable "proxy" se entiende una variable que si bien no es exactamente la que el modelo necesita, si es una variable que se comporta de manera similar. En este caso el PIB de E.U. es una buena proxy, debido a la estrecha relación entre esta economía y la del mundo.

Figura 3.1: PIB real de Colombia con respecto al de Estados Unidos



3.3.1. Creación de las variables dummy

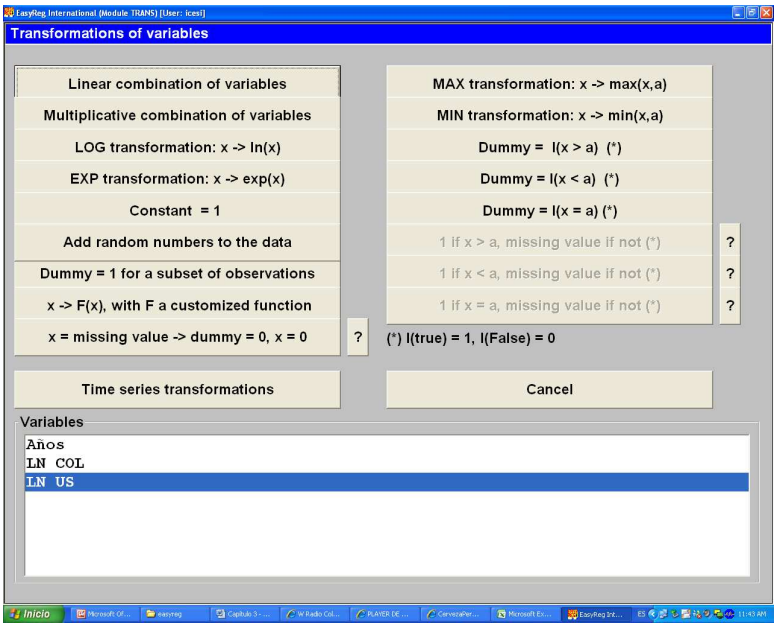
Una vez las series del Logaritmo del PIB de Colombia y de Estados Unidos sean leídas en EasyReg, necesitamos crear una variable dummy. EasyReg provee una variedad de posibilidades para la creación de variables dummy, como por ejemplo: Si queremos que la dummy tome el valor de uno si es mayor, menor o igual que una constante y de cero en el caso contrario.

En este caso queremos simplemente una dummy como la siguiente:

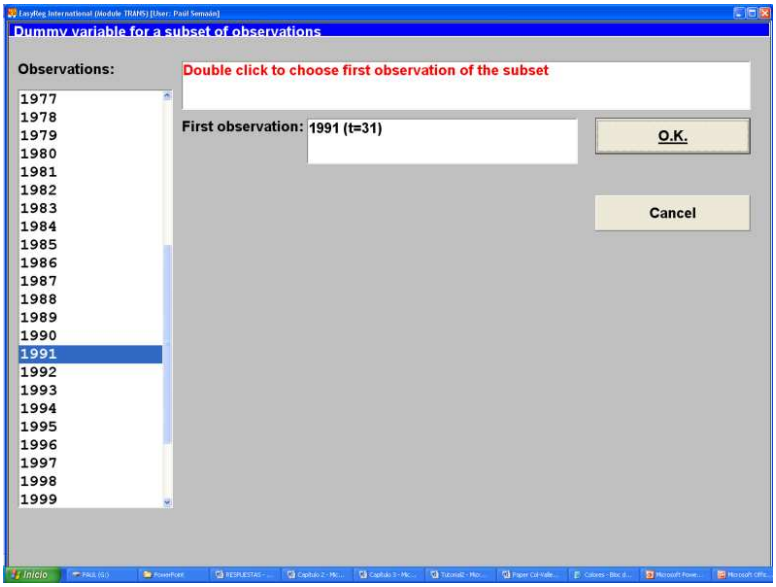
$$D_t = \begin{cases} 1 & si \ t \geq 1991 \\ 0 & o.w. \end{cases} \tag{3.6}$$

Esta variable se puede crear rápidamente con los siguientes pasos. Primero, vaya al menú “Menu/Input/Transform variables”.

En la siguiente ventana, haga clic en el botón "Dummy = 1 for a Subset of Observations".



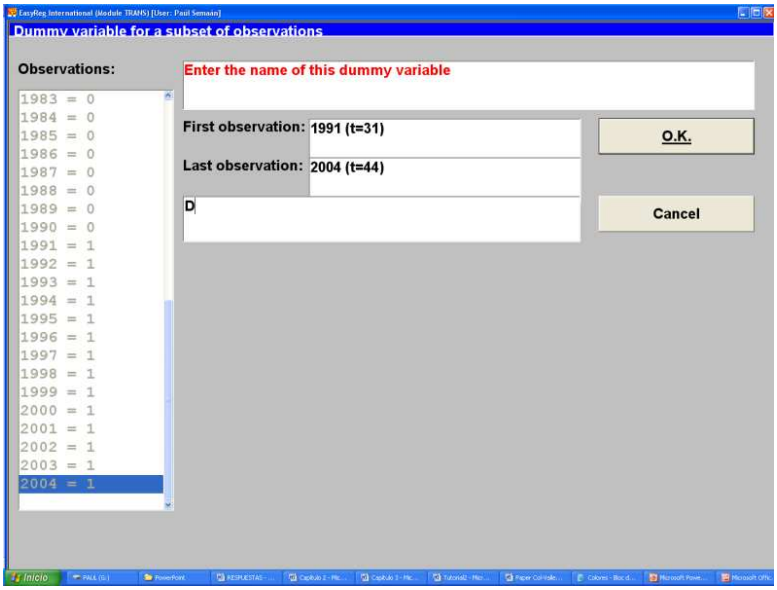
Posteriormente se le pregunta para cuál observación quiere usted que la variable dummy empiece a tomar el valor de uno. Haga doble clic en la ventana de “Observations” y nuevamente en “1991”. Inmediatamente aparecerá un botón en el lado derecho que le pide confirmar su selección. Haga clic en el botón “OK”.



En la siguiente ventana se le pregunta para cuál observación quiere usted que la dummy termine de tomar el valor de uno. Haga doble clic en la ventana de “Observations” en “2004”. Inmediatamente aparecerá un botón en el lado derecho que le pide confirmar su última selección. Haga clic en el botón

"OK".

EasyReg preguntará por el nombre a asignar a la variable dummy. Escriba "D" en el espacio disponible y presione el botón "OK" y nuevamente el botón "OK". ¡La variable dummy ha sido creada!, y ha sido denominada con la letra "D".

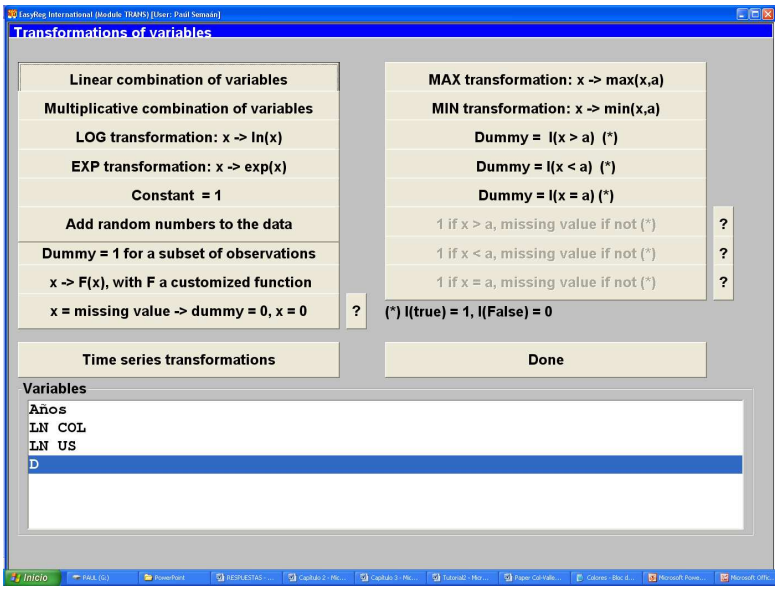


Note que la ventana "Transformations of variables"' permite crear variables dummy para casos como los siguientes:

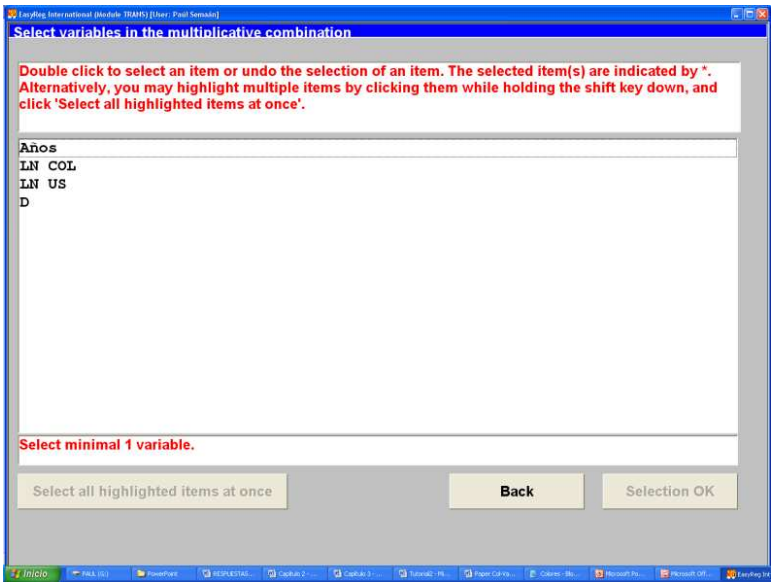
1. $D_t = \begin{cases} 1 & si\ \mathrm{X_i} \geq a \\ 0 & o.w. \end{cases}$ ("Dummy =I(x>a) (*)")

2. $D_t = \begin{cases} 1 & si\ \mathrm{X_i} \leq a \\ 0 & o.w. \end{cases}$ ("Dummy =I(x<a) (*)")

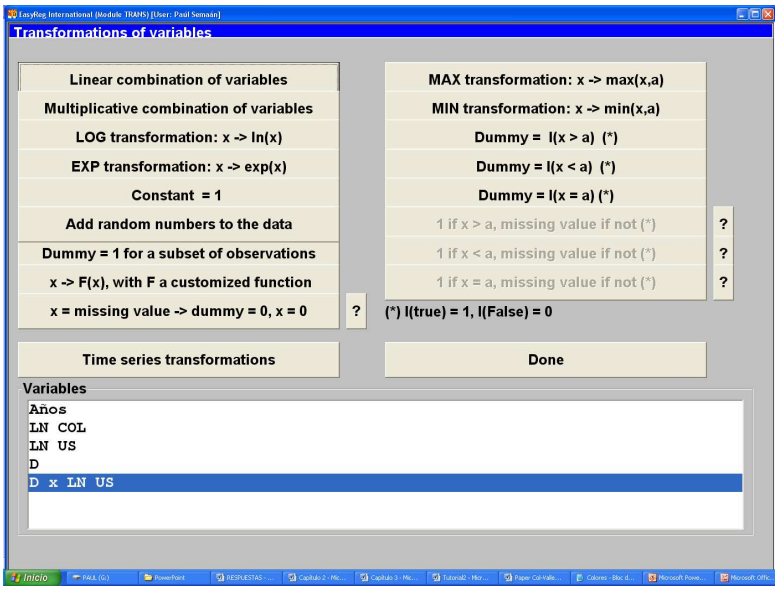
3. $D_t = \begin{cases} 1 & si\ \mathrm{X_i} = a \\ 0 & o.w. \end{cases}$ ("Dummy =I(x=a) (*)").



Ahora necesitamos crear la variable $D_t(f_t)$, en nuestro caso f_t corresponde a $LN(US_t)$. Para esto, haga clic en el botón "Multiplicative combination of variables" de la ventana "Transformations of variables". Aparecerá la siguiente ventana:



Escoja, haciendo doble clic en ellas, las variables D_t, (D) y f_t $(LNUS)$ y posteriormente en el botón "Selection O.K.". Ahora escriba el exponente que le quiere asignar a cada una de las variables escogidas (uno para ambas) y presione "OK" dos veces. Se creará la variable "$D \times LNUS$".



3.3.2. Estimación del modelo

Ahora sí podemos estimar el modelo deseado:

$$Y_t = \beta_0 + \beta_1 f_t + \alpha_2 D_t + \beta_2 (D_t f_t) + \varepsilon_t. \tag{3.7}$$

Para ver los detalles de cómo efectuar la estimación remítase al Capítulo 1. Los resultados de la estimación se presentan en el cuadro 3.1. Asegúrese que puede replicar estos resultados.

Noten que la diferencia en el intercepto, así como la diferencia en la pendiente y el intercepto entre el periodo 1991-2004 y el período previo, son individualmente significativamente diferentes de cero. Ahora sería buena idea hacer una prueba conjunta para ver si $\alpha_2 = \beta_2 = 0$. Como lo discutimos anteriormente, esto se puede hacer fácilmente en EasyReg. El respectivo estadístico de Wald es 84.33[3] (con un p-valor de 0.000). Así, podemos rechazar la hipótesis nula que $\alpha_2 = \beta_2 = 0$. En otras palabras, efectivamente existe un cambio estructural en la relación en la elasticidad del PIB de Colombia con respecto al de Estados Unidos (ver figura 3.2).

[3] Asegúrese que usted puede replicar este resultado

Cuadro 3.1: Estimación del modelo 3.7

	Variable dependiente Y_t Estadísticos t entre paréntesis
	Ecuación 3.7 1961 - 2004 MCO
Constante	-1.4089 (5.108)***
f_t	1.4144 (42.589)***
D_t	6.3366 (6.432)***
$D_t f_t$	-0.7184 (-6.505)***
R^2	0.991
F-Global	1461.99***
No. de obs.	44

(*) nivel de significancia: 10 %

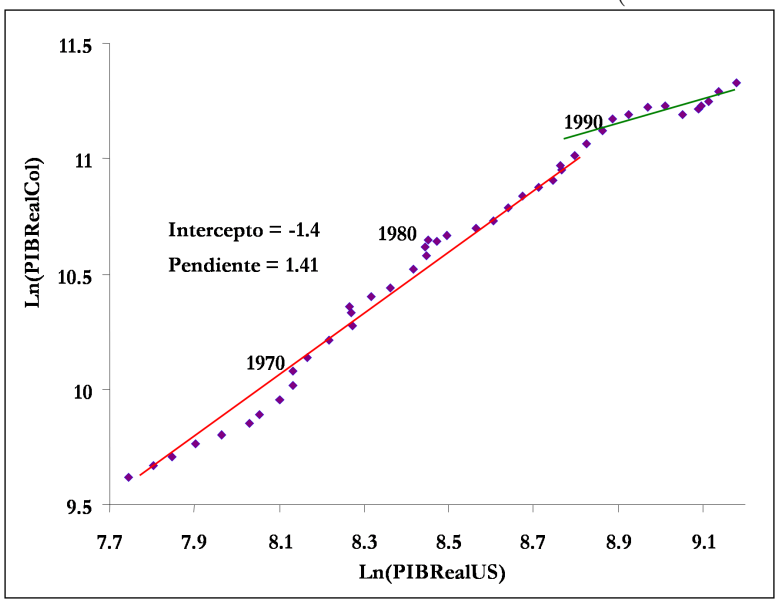
(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos Cuadrados Ordinarios

De los coeficientes que acompañan a las variables f_t y $D_t f_t$ se puede concluir que para el periodo de estudio anterior a la apertura económica $t < 1991$ los crecimientos económicos del resto del mundo generaban un incremento en el crecimiento de la economía colombiana aun mayor, dado que en este caso $\beta_1 > 0$. Por otra parte, en los periodos posteriores al año 1991, en el cual se inicia el proceso de apertura económica de Colombia, se observa que $0<\beta_1 + \beta_2<1$, entonces ante incrementos en el nivel de crecimiento de la economía mundial, la economía doméstica crece en una menor cuantía a este aumento. En la figura 3.2 se ilustran las dos regresiones estimadas.

Figura 3.2: Cambio estructural en la relación del PIB (Colombia-EEUU)



3.4. Ejercicios

Usted como practicante de la Bolsa de Valores de Colombia (BVC), ha notado que se han dado rumores especulativos en el mercado cambiario del peso contra el dólar americano, ocasionando un fuerte aumento en el segundo trimestre del año, notándose especialmente en los días de la semana que corresponden a la apertura y cierre del mercado. Su jefe le ha encomendado como tarea investigar si la teoría del mercado eficiente se cumple en los diferentes días de la semana, para tratar de descubrir un patrón en el comportamiento de los especuladores.

En concreto, el "efecto día de la semana" (DOW) en el rendimiento de la moneda extranjera permite tomar ventaja de los patrones de comportamiento de los mercados diseñando estrategias de negociación (ver Alonso y Romero (2008)).

La presencia de un patrón de comportamiento en los retornos al interior de la semana podría indicar la existencia de retornos condicionales según el día de la semana ofreciendo así oportunidades de arbitraje al inversor.

La existencia de estas reglas de negociación, válidas a largo plazo, implicarían la violación de la hipótesis de los mercados financieros eficientes. En general si existen patrones predeterminados, previsibles y disponibles abiertamente que permitan generar utilidades, entonces existirá evidencia en contra de la EMH.

1. Utilice los datos del archivo dummy.xls, en el cual encontrará la Tasa Representativa del Mercado diaria (peso colombiano por cada dólar estadounidense), para determinar la serie de rendimientos a partir del primer día de negociación del año 2005 hasta el último día de negociación de abril de 2006, grafique la serie de rendimientos y comente sobre el valor esperado de los rendimientos que revela dicho gráfico.

2. Con el fin de dar respuesta a su jefe usted debe crear un modelo que muestre los efectos que genera cada uno de los días de la semana sobre el rendimiento promedio de la TRM, para conocer

el "efecto del día de la semana" (DOW). Escriba el modelo a estimar y demuestre que este modelo sí sirve para este efecto.

3. Estime el modelo y reporte los resultados en una Tabla.

4. Continuando con la pregunta anterior, realice las pruebas de significancia conjunta e individual que crea pertinentes.

5. A partir de estos resultados estime el modelo que usted considere más adecuado, resuma sus resultados en una tabla.

6. Interprete el significado de los coeficientes estimados, comente sobre el nivel de significancia y el signo esperado.

7. A partir de los resultados anteriores ¿qué medidas puede tomar usted para sacar ventaja en el mercado cambiario? (Describa por completo el procedimiento realizado); ¿se puede concluir si el mercado cambiario es eficiente?

Parte II

Problemas econométricos

CAPÍTULO 4

Multicolinealidad

Objetivos del capítulo

Este capítulo tiene como objetivo ilustrarle al lector cómo es posible:

- Identificar, mediante el uso de EasyReg los diferentes síntomas que presenta un modelo estimado en presencia de multicolinealidad.
- Efectuar con EasyReg diferentes pruebas formales, con el fin de detectar multicolinealidad en el modelo. Estas pruebas son i) El cálculo de las correlaciones entre los coeficientes asociados a las pendientes y al intercepto, ii) El cálculo de la matriz de correlaciones de las variables explicativas y iii) La prueba de Belsley, Kuh y Welsh (1980)

4.1. Introducción

Como lo habíamos discutido en capítulos anteriores, si el modelo de regresión múltiple cumple con los supuestos que se resumen en el recuadro 4.1, entonces el Teorema de Gauss-Markov demuestra que los estimadores MCO son MELI (Mejor Estimador Lineal Insesgado) y por lo tanto tienen la menor varianza posible cuando se comparan con todos los estimadores lineales posibles.

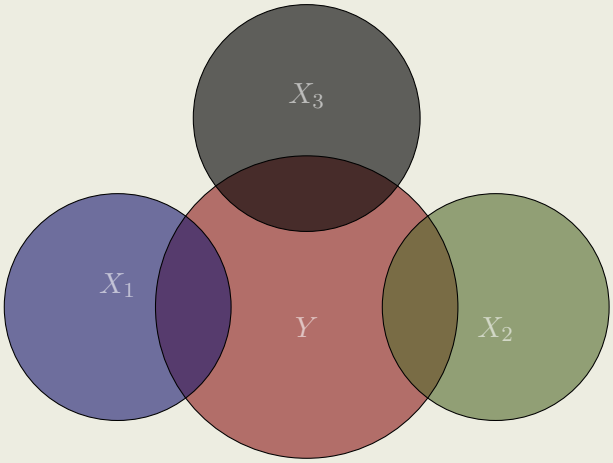
Recuadro 4.1 Supuestos del modelo de regresión múltiple

1. Relación lineal entre $\mathbf{y}$ y $X_2, X_3, ..., X_k$
2. Las $X_2, X_3, ..., X_k$ son fijas y linealmente independientes (i.e. la matriz X tiene rango completo)
3. el vector de errores ε satisface:
   - Media cero ($E[\varepsilon] = 0$),
   - Varianza constante
   - No autocorrelación
     Es decir: $\varepsilon_i \sim i.i.d\left(0, \sigma^2\right)$ ó $\varepsilon_{n\times 1} \sim \left(0_{n\times 1}, \sigma^2 I_n\right)$

Ahora veamos qué ocurre si se viola una parte del supuesto 2, en especial, las variables explicativas (X´s) no sean linealmente independientes entre sí. Es importante anotar que la violación de la otra parte del supuesto no tiene grandes implicaciones sobre el resultado que los estimadores MCO sean MELI (ver el capítulo 2 de Greene (2003)). La violación de este supuesto se puede entender gráficamente con el ejemplo 4.1

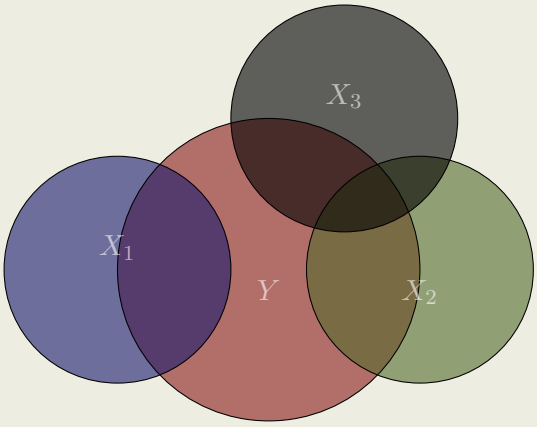
Ejemplo 4.1 Relación entre las variables

Los supuestos 1 y 2 del teorema del Gauss Markov señalan la existencia de una relación lineal entre las variables explicativas y la dependiente, además de una relación linealmente independiente entre las variables explicativas (X´s). El siguiente gráfico representa estos supuestos. Podemos observar que existe una relación entre la variable explicada y las independientes, pero a su vez independencia entre estas últimas. Los círculos representan las variables y sus intersecciones la relación entre ellas.

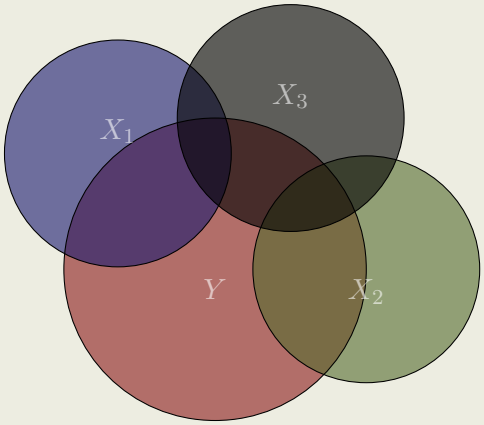


El problema aparece cuando tenemos algún tipo de relación lineal entre las variables independientes o entre un subconjunto de ellas.

En el siguiente gráfico podemos apreciar la presencia de una relación entre las variables X_2 y X_3, es decir, una parte de la información proporcionada por X_2 está a su vez contenida en la variable X_3. En este caso, si X_2 cambia, esto provoca un cambio directo en Y (representado por β_2) y también un cambio indirecto; pues al cambiar X_2, X_3 cambia y esto a su vez provoca el cambio en Y.



O podría darse el caso en el que todas las variables independientes se encuentren relacionadas, es decir todas ellas comparten una parte de la información contenida en cada una.

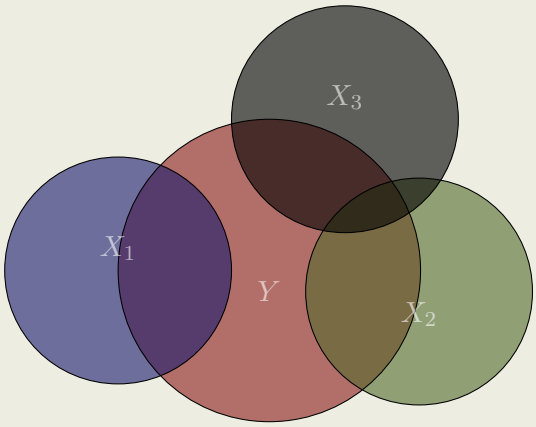


4.2. Los diferentes grados de multicolinealidad

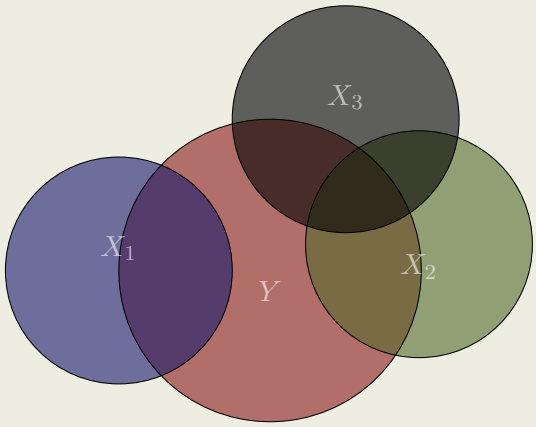
En general, cuando hay cierto grado de relación lineal entre las variable independientes decimos que existe multicolinealidad (o colinealidad). En la práctica, tendremos diferentes grados de multicolinealidad; en el Ejemplo 2 se presentan los cuatro posibles grados o tipos de multicolinealidad.

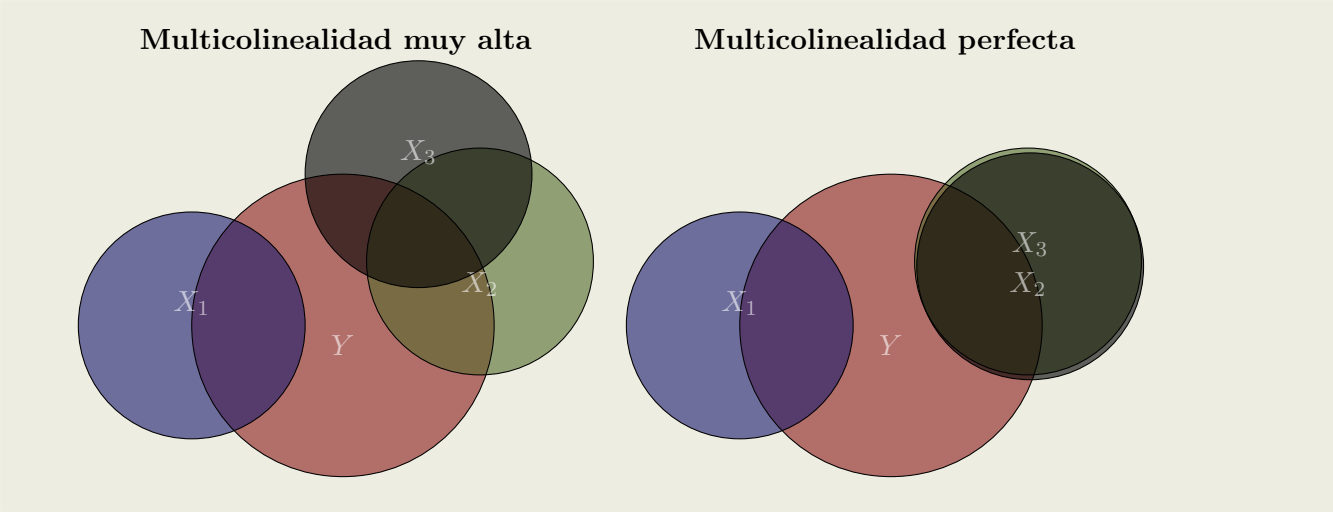
Ejemplo 4.2 Grados de multicolinealidad

Multicolinealidad moderada



Multicolinealidad alta





Discutiremos primero la multicolinealidad perfecta y sus efectos.

4.2.1. Multicolinealidad perfecta

Partamos de un ejemplo. Supongamos que queremos explicar la relación entre el peso en kilogramos de un individuo (kg_i) con las horas diarias promedio de actividad física y las calorías consumidas. Veamos qué sucede si empleamos el siguiente modelo:

$$kg_i = \beta_0 + \beta_1 ha_i + \beta_2 cd_i + \beta_3 cs_i + e_i$$

donde las variables ha_i, cd_i y cs_i corresponden a los promedios de horas diarias de actividad física, calorías consumidas por día y calorías consumidas por semana por el individuo i, respectivamente.

Como podemos apreciar, las variables cd_i y cs_i presentan una relación lineal perfecta (y por tanto multicolinealidad perfecta) debido a que siempre vamos a tener que $7cd_i = cs_i$ y por lo tanto las X's no son linealmente independientes entre sí.

Matricialmente este hecho se puede representar de la siguiente manera:

$$\mathbf{y} = \begin{bmatrix} Kg_1 \\ Kg_2 \\ \vdots \\ Kg_n \end{bmatrix}_{n\times 1} \qquad \mathbf{X} = \begin{bmatrix} 1 & ha_1 & cd_1 & 7cd_1 \\ 1 & ha_2 & cd_2 & 7cd_2 \\ 1 & ha_3 & cd_3 & 7cd_3 \\ 1 & ha_4 & cd_4 & 7cd_4 \\ 1 & ha_5 & cd_5 & 7cd_5 \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{n\times 4} \tag{4.1}$$

Las consecuencias de esta relación entre dos columnas de la matriz $\mathbf{X}$ es que ésta no tendrá rango columna completo y por tanto $\mathbf{X^T X}$ no tendrá rango completo, es decir $\det(\mathbf{X^T X}) = 0$. Así, $\left(\mathbf{X^T X}\right)^{-1}$ no existirá y por tanto $\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$ no existirá. En este caso, si queremos eliminar la multicolinealidad perfecta entre cd_i y cs_i, debemos eliminar cualquiera de las dos variables, ya que ambas están aportando la misma información al modelo.

En resumen, el problema que se presenta en presencia de multicolinealidad perfecta es que las columnas de la matriz de las $X's$ no son linealmente independientes y esto implicará que el estimador de MCO no existirá. Por lo tanto, si un modelo presenta multicolinealidad perfecta, entonces EasyReg o cualquier otro paquete estadístico reportará un error y no podrá estimar los coeficientes. Finalmente, aunque es imposible estimar un modelo con este problema, de todas formas por el mismo diseño del modelo es fácil de detectar y de resolver el problema.

También es posible estar expuesto a la presencia de multicolinealidad perfecta al utilizar variables dummy. Por ejemplo, supongamos que queremos ver el efecto del sector de la economía donde se emplea un individuo sobre el salario (w_i). Supongamos que el modelo, sin tener en cuenta el efecto del sector económico, es:

$$w_i = \lambda_1 + \lambda_2 (E_i) + \lambda_3 (C_i) + \mu_i \tag{4.2}$$

donde E_i y C_i representan los años de educación y los años de capacitación del individuo i respectivamente

Ahora, supongamos que la economía tiene tres diferentes sectores: primario (Agricultura, minería, ganadería, etc.), secundario (Manufacturas, etc.), terciario (Comercio, servicios, etc.) y además un individuo solo puede recibir un salario si trabaja en uno de esos tres sectores.

Esto implica generar las siguientes variables dummy:

$$D_{1i} = \begin{cases} 1 & \text{si i} \in \text{Sec. Primario} \\ 0 & \text{o.w.} \end{cases}$$

$$D_{2i} = \begin{cases} 1 & \text{si i} \in \text{Sec. Secundario} \\ 0 & \text{o.w.} \end{cases}$$

$$D_{3i} = \begin{cases} 1 & \text{si i} \in \text{Sec. Terciario} \\ 0 & \text{o.w.} \end{cases}$$

entonces, nuestro modelo se convierte en:[1]

$$\begin{aligned} w_i &= \lambda_1 + \lambda_2 (E_i) + \lambda_3 (C_i) + \lambda_4 D_{1i} + \lambda_5 D_{2i} + \lambda_6 D_{3i} + \mu_i \\ E[w_i] &= \lambda_1 + \lambda_2 E_i + \lambda_3 C_i + \lambda_4 E[D_{1i}] + \lambda_5 E[D_{2i}] + \lambda_6 E[D_{3i}] \end{aligned}$$

Pero si analizamos el anterior modelo mediante su expresión matricial, nos damos cuenta rápidamente del problema que aparece. Matricialmente el modelo será:

$$\mathbf{y} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}_{n\times 1} \qquad \mathbf{X} = \begin{bmatrix} 1 & E_1 & C_1 & 1 & 0 & 0 \\ 1 & E_2 & C_2 & 0 & 0 & 1 \\ 1 & E_3 & C_3 & 0 & 1 & 0 \\ 1 & E_4 & C_4 & 0 & 0 & 1 \\ 1 & E_5 & C_5 & 0 & 0 & 1 \\ 1 & E_6 & C_6 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}_{n\times 6}$$

[1] Por simplicidad solo consideraremos cambios en el intercepto.

Para todas las observaciones ($\forall_i$), tenemos que $D_{1i} + D_{2i} + D_{3i} = 1$. Además, la columna de la constante será igual a 1 en cada observación. Por lo anterior, concluimos que las $X's$ no son linealmente independientes, al existir una combinación lineal de las columnas 4, 5 y 6 que es exactamente igual a la primera columna.

Si queremos no tener multicolinealidad perfecta entre las variables dummy y el intercepto debemos eliminar una de las dummy obteniendo el siguiente modelo:

$$w_i = \lambda_1 + \lambda_2 (E_i) + \lambda_3 (C_i) + \lambda_4 D_{1i} + \lambda_5 D_{2i} + \mu_i$$

Ahora,

$$D_{1i} + D_{2i} \neq 1 \quad \forall i$$

y por tanto, no hay relación lineal entre las dummy y el intercepto.

En este caso $E[w_i] = \lambda_1 + \lambda_2 E_i + \lambda_3 C_i + \lambda_4 E[D_{1i}] + \lambda_5 E[D_{2i}]$. Es decir:

$$E[w_i] = \begin{cases} (\lambda_1 + \lambda_4) + \lambda_2 E_i + \lambda_3 C_i & si \ \ i \in \text{sec .Prim} \\ (\lambda_1 + \lambda_5) + \lambda_2 E_i + \lambda_3 C_i & si \ \ i \in \text{sec .Sec} \\ \lambda_1 + \lambda_2 E_i + \lambda_3 C_i & si \ \ i \in \text{sec .Ter} \end{cases}$$

En general cuando usamos variables dummy tenemos que tener cuidado si existen j posibilidades diferentes entonces usamos $j-1$ variables dummy ó j variables dummy y quitamos el intercepto, así evitamos multicolinealidad perfecta.

4.2.2. Consecuencias de la multicolinealidad no perfecta

En la práctica, el problema de multicolinealidad perfecta es muy raro, pero sí es común contar con modelos con variables independientes altamente correlacionadas entre sí, sin que esta relación sea perfecta. Supongamos que dos o más variables están relacionadas (pero no de manera perfecta), en este caso se tendrá que el $\det(\mathbf{X^T X}) = |\mathbf{X^T X}|$ existe, pero tiende a cero. Por lo tanto, la matriz $(\mathbf{X^T X})^{-1}$ sí existe pero en general tendrá valores muy grandes. Recuerde que $(\mathbf{X^T X})^{-1} = \frac{1}{|\mathbf{X^T X}|} Adj(\mathbf{X^T X})$.

Como la matriz de varianzas y covarianzas de los coeficientes estimados depende de la matriz[2] $(\mathbf{X^T X})^{-1}$, entonces las varianzas de los $\beta's$ tenderán a ser a su vez muy grandes. Esto implica que los t-calculados de los $\beta's$ para probar la significancia de los coeficientes estimados sean relativamente bajos,[3] y así se tenderá a no rechazar la hipótesis nula de no significancia individual de los coeficientes. Así mismo una matriz $(\mathbf{X^T X})^{-1}$ con valores relativamente grandes, implicará que la suma cuadrada de la regresión ($SSR = \left((\mathbf{X^T X})^{-1}\mathbf{X^T y}\right)^T \mathbf{X^T y} - n\bar{\mathbf{y}}^2$) será relativamente grande y por tanto el R^2 y el $F - Global$ serán relativamente grandes ya que estos dependen del SSR.

Asi, los síntomas más comunes de la multicolinealidad no perfecta se pueden resumir de la siguiente manera:

1. t calculados bajos acompañados de $F - Global$ y R^2 altos

[2] $Var\left[\hat{\beta}\right] = \sigma^2 (\mathbf{X^T X})^{-1}$.

[3] Recordemos que $t_c = \frac{\hat{\beta}_i}{s_{\hat{\beta}_i}}$.

2. Sensibilidad de los $\beta's$ estimados a cambios pequeños en la muestra[4]

3. Sensibilidad de los $\beta's$ a la inclusión o exclusión de regresores[5]

No obstante la multicolinealidad no perfecta provoca estos síntomas, en muchos casos este problema es ignorado por los econometristas. La razón es que este es un problema de los datos y no un problema del modelo o del método de estimación.

Existen algunas técnicas estadísticas que permiten lidiar con el problema de la multicolinealidad no perfecta, como por ejemplo la "ridge regression" o el método de componentes principales para condensar la información de las variables relacionadas en una sola. Pero en la realidad las prácticas más empleadas para "lidiar" con la multicolinealidad no perfecta son:

- Descartar una variable de las que presentan el problema. Esta aproximación no será aceptable para el investigador si la especificación del modelo inicial proviene directamente de la teoría económica. La omisión de una variable relevante (según la teoría) para explicar la variable dependiente provoca un sesgo de omisión. Problema que es más grave que el problema causado por la multicolinealidad no perfecta. Así, este remedio puede ser más costoso que el mismo problema de multicolinealidad.

- Transformar las variables relacionadas, de tal forma que el modelo involucre razones de las variables y no las variables en sus niveles. Esta aproximación también puede no ser aceptable, pues la teoría económica sugiere los cambios de los niveles de la variable dependiente, dados cambios en los niveles de las variables explicativas y no en las razones de estas variables.

- Aumentar la muestra. Al aumentar la muestra, se puede brindar más información que permita mejorar la precisión en la estimación de los coeficientes y por tanto la disminución del error estándar de estos. El problema con esta aproximación es que en la mayoría de los casos el investigador emplea la mayor cantidad de datos disponibles para el problema y aumentar la muestra es una misión casi imposible.

- No hacer nada. Si bien se identifica la presencia de éste problema, en muchos casos no se realiza algún procedimiento para resolverlos, pues la multicolinealidad no perfecta no afecta las propiedades de MELI de los MCO.[6] El problema aparece en la interpretación de los coeficientes y en algunos casos en la reducción de los t-calculados que podrían ser lo suficientemente grandes como para rechazar la nula de no significancia si la multicolinealidad no perfecta no estuviese presente. Así, si esta es la opción que se adopta para "lidiar" con la multicolinealidad, se tendrá que ser muy cuidadoso con la interpretación de los coeficientes y con la inferencia respecto a estos.

En conclusión, de encontrarse multicolinealidad no perfecta en un modelo, entonces **sólo la teoría económica** nos podrá servir para decidir si eliminar o no una variable.

Por otro lado, si el objetivo del modelo estimado con multicolinealidad es la de producir pronósticos, entonces el problema de multicolinealidad no es muy importante, siempre y cuando la relación entre las variables explicativas se mantenga en el período proyectado.

[4]Esto ocurre porque la matriz $\left(\mathbf{X^T X}\right)^{-1}$ cambiará mucho con incluir o eliminar una fila de $\mathbf{X}$.

[5]Esto ocurre porque $\left(\mathbf{X^T X}\right)^{-1}$ cambiará mucho con incluir o eliminar una columna de $\mathbf{X}$.

[6]En el Apéndice 3 del Capítulo 1, se presenta la demostración del teorema de Gauss-Markov. En esta demostración, es fácil notar que las propiedades de insesgadez y mínima varianza de los estimadores MCO dependen de los supuestos asociados al término de error y no a qué tan grande o pequeño sea el determinante de la matriz $X^T X$.

4.3. Pruebas para la detección de multicolinealidad

Además de chequear los síntomas, en la práctica es necesaria la utilización de pruebas más formales con el fin de detectar la presencia de multicolinealidad no perfecta. A continuación se describen tres de las pruebas más utilizadas para este fin.

4.3.1. Prueba del determinante de la matriz de correlaciones de las X´s

Esta primera prueba consiste en verificar el comportamiento de la matriz de correlaciones parciales entre las $X's$. En especial definamos RX como la matriz de correlaciones de las $X's$, es decir:

$$RX = \begin{bmatrix} 1 & \hat{\rho}_{X_2,X_3} & \cdots & \hat{\rho}_{X_2,X_k} \\ & 1 & \cdots & \hat{\rho}_{X_3,X_k} \\ & & \ddots & \vdots \\ & & & 1 \end{bmatrix} \tag{4.3}$$

Entonces será interesante conocer el comportamiento del determinante de la matriz de correlaciones ($\det(RX) = |RX|$). Se puede demostrar fácilmente que:

- La multicolinealidad es perfecta si el $|RX| = 0$.
- Si el $|RX| = 1$, entonces no existe problema de multicolinealidad.
- Y finalmente si el $|RX| \to 0$, entonces el problema de multicolinealidad es grave.

Es importante recordar que el determinante de una matriz se puede calcular fácilmente multiplicando sus valores propios; es decir

$$|RX| = \prod_{i=1}^{k-1} \lambda_i$$

donde λ_i son los valores propios (Eigenvalues) de la matriz RX.

4.3.2. Prueba de Belsley, Kuh y Welsh (1980)

Esta prueba, también conocida con el nombre de prueba Kappa, se basa en los valores propios de la matriz de correlaciones de las X´s. Empleando el valor máximo y mínimo de estos se construye la razón $k(X)$ de la siguiente manera:

$$\kappa(X) = \frac{\sqrt{\lambda_{Max}}}{\sqrt{\lambda_{Min}}} \tag{4.4}$$

Belsley et al. (1980) concluyen que:

- Si el valor $\kappa(X) > 30$, entonces la multicolinealidad es perfecta
- Si se encuentra entre $20 < \kappa(X) < 30$, entonces el problema de multicolinealidad es grave.
- Y finalmente si $k = 1$, entonces no habrá problema de multicolinealidad.

4.3.3. Prueba de correlaciones entre los $\beta's$ estimados

Esta última aproximación se basa en la matriz de varianzas y covarianzas de los $\beta's$ estimados. La correlación entre los coeficientes se obtiene de la siguiente fórmula:

$$\hat{\rho}_{\hat{\beta}_1,\hat{\beta}_2} = corr\left(\hat{\beta}_1, \hat{\beta}_2\right) = \frac{\widehat{Cov\left(\hat{\beta}_1, \hat{\beta}_2\right)}}{s_{\hat{\beta}_1} s_{\hat{\beta}_2}} \tag{4.5}$$

El problema de multicolinealidad será grave si el valor absoluto de la correlación entre dos coeficientes es mayor a 0.8. A continuación se presenta un ejemplo en el que se puede apreciar la aplicación de cada una de estas pruebas.

4.4. Análisis del efecto discriminatorio de género en las diferencias salariales en Colombia

El objetivo de este ejercicio es aplicar las tres pruebas de multicolinealidad anteriormente descritas y de ser necesario solucionar dicho problema. Para este fin y con la misma metodología seguida por Bernat (2004) se analizan las brechas salariales entre hombres y mujeres en el municipio de Santiago de Cali en el año 2003.

Uno de los métodos más utilizados para determinar el efecto de la discriminación en la brecha salarial es la propuesta por Oaxaca (1973), quienes sugieren una técnica para medir la diferencia que tiene sobre el ingreso, características como el capital humano entre géneros. Para estimar este efecto, se emplea un modelo de ecuaciones Mincerianas (Mincer (1974)) que fueron pensadas para estimar la rentabilidad de la educación. Aunque originalmente el autor propone estimar dos ecuaciones por separado, una para hombres y otra para mujeres, en nuestro caso sólo estimaremos una ecuación y añadiremos una variable dummy para capturar el efecto que tiene el género sobre el salario.

De esta manera, el modelo a estimar es el siguiente:

$$\begin{aligned} \ln(ih_i) =& \beta_0 + \beta_1 yedu_i + \beta_2 exp_i + \beta_3 exp_i^2 + \beta_4 D_i \\ &+ \beta_5 Dyedu_i + \beta_6 Dexp_i + \beta_7 Dexp_i^2 + \varepsilon_i \end{aligned} \tag{4.6}$$

donde

$$D_i = \begin{cases} 1 & \text{si el individuo } i \text{ es hombre} \\ 0 & o.w. \end{cases}$$

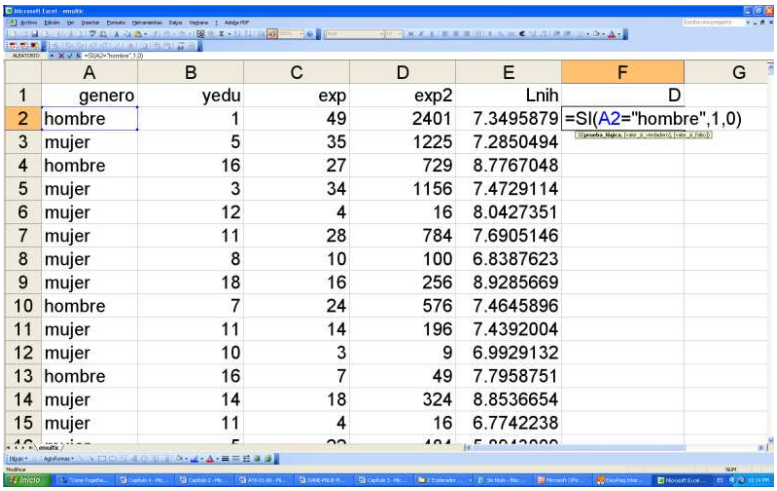
$\ln(ih_i)$ representa el logaritmo natural del ingreso por hora del individuo i, $yedu_i$ y exp_i denotan los años de educación y de experiencia del individuo i. Los datos se encuentran en el archivo emulti.xls. Antes de analizar los datos, es claro que la especificación de este modelo incluye dos variables que están relacionadas, pero no de manera lineal; es decir, exp y exp^2. Como la relación es cuadrática y no lineal, no hay razón por la cual esperar, a priori, la presencia de multicolinealidad perfecta o no perfecta.

4.4.1. Creación de variables dummy en Excel

En este caso no es posible crear la variable dummy en EasyReg,[7] por lo tanto es necesario que la creemos directamente en Excel. Esto lo podemos hacer fácilmente mediante un condicional. En primer

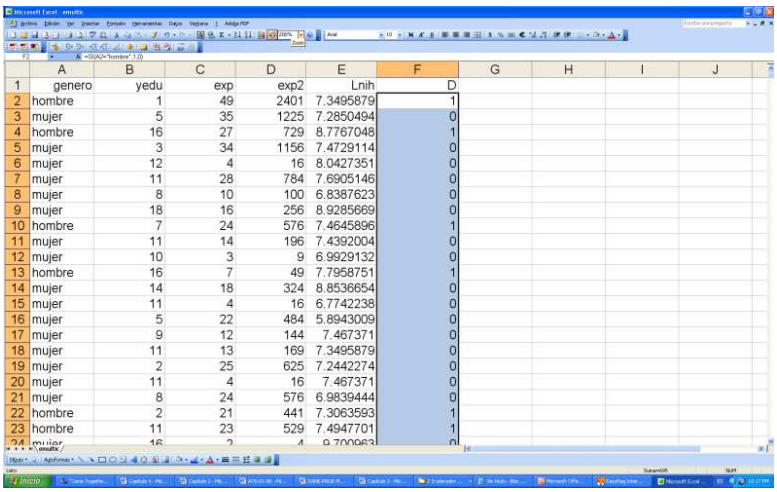
[7] En este caso los datos tienen una columna que contiene caracteres de letras y eso no permite que EasyReg lea los datos.

lugar abra los datos, observará que la primera columna corresponde al género, a continuación ubíquese en la columna F, la cual corresponderá a la variable dummy(D). Como queremos que la variable tome el valor de 1 si el género es hombre y 0 de lo contrario, entonces empleemos el siguiente condicional: =Si(A2="hombre", 1, 0)



	A	B	C	D	E	F	G
1	genero	yedu	exp	exp2	Lnih	D	
2	hombre	1	49	2401	7.3495879	=SI(A2="hombre",1,0)	
3	mujer	5	35	1225	7.2850494		
4	hombre	16	27	729	8.7767048		
5	mujer	3	34	1156	7.4729114		
6	mujer	12	4	16	8.0427351		
7	mujer	11	28	784	7.6905146		
8	mujer	8	10	100	6.8387623		
9	mujer	18	16	256	8.9285669		
10	hombre	7	24	576	7.4645896		
11	mujer	11	14	196	7.4392004		
12	mujer	10	3	9	6.9929132		
13	hombre	16	7	49	7.7958751		
14	mujer	14	18	324	8.8536654		
15	mujer	11	4	16	6.7742238		

Posteriormente, hacemos doble clic sobre el extremo inferior derecho de la celda en que escribimos el condicional y que ahora posee el valor de 1, se arrastrará automáticamente la fórmula hasta el final de la columna. La variable dummy ya ha sido creada.



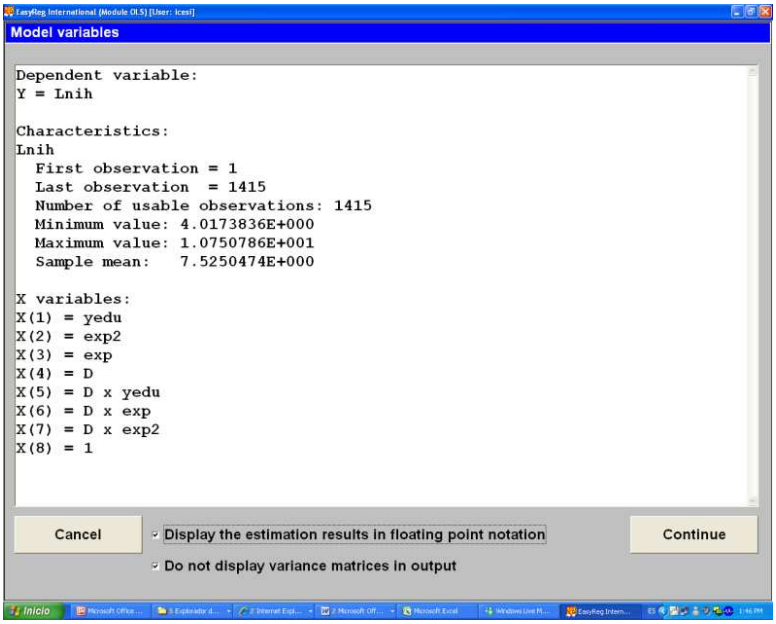
	A	B	C	D	E	F	G	H	I	J
1	genero	yedu	exp	exp2	Lnih	D				
2	hombre	1	49	2401	7.3495879	1				
3	mujer	5	35	1225	7.2850494	0				
4	hombre	16	27	729	8.7767048	1				
5	mujer	3	34	1156	7.4729114	0				
6	mujer	12	4	16	8.0427351	0				
7	mujer	11	28	784	7.6905146	0				
8	mujer	8	10	100	6.8387623	0				
9	mujer	18	16	256	8.9285669	0				
10	hombre	7	24	576	7.4645896	1				
11	mujer	11	14	196	7.4392004	0				
12	mujer	10	3	9	6.9929132	0				
13	hombre	16	7	49	7.7958751	1				
14	mujer	14	18	324	8.8536654	0				
15	mujer	11	4	16	6.7742238	0				
16	mujer	5	22	484	5.8943009	0				
17	mujer	9	12	144	7.467371	0				
18	mujer	11	13	169	7.3495879	0				
19	mujer	2	25	625	7.2442274	0				
20	mujer	11	4	16	7.467371	0				
21	mujer	8	24	576	6.9839444	0				
22	hombre	2	21	441	7.3063593	1				
23	hombre	11	23	529	7.4947701	1				

4.4.2. Pruebas de multicolinealidad

En primer lugar cargue los datos (no olvide eliminar la columna de género debido a que posee texto), cree las variables restantes ($Dyedu_i$, $Dexp_i$ y $Dexp_i^2$) en EasyReg mediante la transformación de variables.[8] A continuación estime el modelo por el método de mínimos cuadrados ordinarios (MCO).[9] Deshabilite la opción "Do not display variance matrices in output", de tal manera que EasyReg muestra la matriz de varianzas y covarianzas de los $\beta's$ estimados.

[8]Si necesita ayuda en este procedimiento, remítase al capítulo 3.

[9]Para ver detalles de cómo cargar los datos o estimar el modelo, remítase al capítulo 1.



La salida de los resultados obtenidos por EasyReg se muestra a continuación:

Recuadro 4.2 Output de EasyReg modelo 4.4.2

Dependent variable:
Y = Lnih
Characteristics:
Lnih
First observation = 1
Last observation = 1415
Number of usable observations: 1415
Minimum value: 4.0173836E+000
Maximum value: 1.0750786E+001
Sample mean: 7.5250474E+000

X variables:
X(1) = yedu
X(2) = exp
X(3) = exp2
X(4) = D
X(5) = D x yedu
X(6) = D x exp
X(7) = D x exp2
X(8) = 1

Model:
Y = b(1)X(1) +.....+ b(8)X(8) + U,
where U is the error term, satisfying
E[U|X(1),...,X(8)] = 0.
OLS estimation results

Parameters	Estimate	t-value (S.E.) [p-value]	H.C. t-value (H.C. S.E.) [H.C. p-value]
b(1)	0.13653	23.743 (0.00575) [0.00000]	21.688 (0.00629) [0.00000]

b(2)	0.03048	5.407	5.210
		(0.00564)	(0.00585)
		[0.00000]	[0.00000]
b(3)	-0.00023	-1.990	-2.041
		(0.00012)	(0.00011)
		[0.04660]	[0.04123]
b(4)	0.24800	1.983	1.709
		(0.12504)	(0.14508)
		[0.04733]	[0.08737]
b(5)	-0.01131	-1.420	-1.257
		(0.00797)	(0.00900)
		[0.15572]	[0.20867]
b(6)	0.00339	0.433	0.383
		(0.00784)	(0.00886)
		[0.66519]	[0.70152]
b(7)	-0.00008	-0.495	-0.452
		(0.00015)	(0.00017)
		[0.62086]	[0.65136]
b(8)	5.66732	62.633	55.859
		(0.09048)	(0.10146)
		[0.00000]	[0.00000]

Notes:
1: S.E. = Standar* error
2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
3: The two-sided p-values are based on the normal approximation.

Effective sample size (n): 1415
Variance of the residuals: 0.329689
Standard error of the residuals (SER): 0.574185
Residual sum of squares (RSS): 463.872119
(Also called SSR = Sum of Squared Residuals)
Total sum of squares (TSS): 851.238679
R-square: 0.4551
Adjusted R-square: 0.4524

Overall F test: F(7,1407) = 167.85
p-value = 0.00000

Significance levels:	10 %	5 %
Critical values:	1.72	2.01
Conclusions:	reject	reject

If the model is correctly specified, in the sense that the conditional expectation of the model error U relative to the X variables equals zero, then the OLS parameter estimators b(1),..,b(8), minus their true values, times the square root of the sample size n, are (asymptotically) jointly normally distributed with zero mean vector and variance matrix:

4.68E-02	3.76E-03	9.71E-05	5.70E-01	-4.68E-02	-3.76E-03	-9.71E-05	-5.70E-01
3.76E-03	4.50E-02	-8.63E-04	4.44E-01	-3.76E-03	-4.50E-02	8.63E-04	-4.44E-01
9.71E-05	-8.63E-04	1.90E-05	-5.68E-03	-9.71E-05	8.63E-04	-1.90E-05	5.68E-03
5.70E-01	4.44E-01	-5.68E-03	2.21E+01	-1.04E+00	-8.72E-01	1.14E-02	-1.16E+01
-4.68E-02	-3.76E-03	-9.71E-05	-1.04E+00	8.99E-02	4.32E-03	1.87E-04	5.70E-01
-3.76E-03	-4.50E-02	8.63E-04	-8.72E-01	4.32E-03	8.71E-02	-1.61E-03	4.44E-01
-9.71E-05	8.63E-04	-1.90E-05	1.14E-02	1.87E-04	-1.61E-03	3.38E-05	-5.68E-03
-5.70E-01	-4.44E-01	5.68E-03	-1.16E+01	5.70E-01	4.44E-01	-5.68E-03	1.16E+01

provided that the conditional variance of the model error U is constant
(U is homoskedastic), or

5.61E-02	9.93E-03	2.27E-05	7.19E-01	-5.61E-02	-9.93E-03	-2.27E-05	-7.19E-01
9.93E-03	4.84E-02	-8.72E-04	5.77E-01	-9.93E-03	-4.84E-02	8.72E-04	-5.77E-01
2.27E-05	-8.72E-04	1.80E-05	-7.34E-03	-2.27E-05	8.72E-04	-1.80E-05	7.34E-03
7.19E-01	5.77E-01	-7.34E-03	2.98E+01	-1.37E+00	-1.29E+00	1.75E-02	-1.46E+01
-5.61E-02	-9.93E-03	-2.27E-05	-1.37E+00	1.15E-01	1.54E-02	6.48E-05	7.19E-01
-9.93E-03	-4.84E-02	8.72E-04	-1.29E+00	1.54E-02	1.11E-01	-1.99E-03	5.77E-01
-2.27E-05	8.72E-04	-1.80E-05	1.75E-02	6.48E-05	-1.99E-03	4.06E-05	-7.34E-03
-7.19E-01	-5.77E-01	7.34E-03	-1.46E+01	7.19E-01	5.77E-01	-7.34E-03	1.46E+01

if the conditional variance of the model error U is not constant
(U is heteroskedastic).

Antes de entrar a calcular los estadísticos para determinar la presencia de multicolinealidad, es importante anotar que los síntomas no están presentes.

Si observamos la estimación realizada por EasyReg, nos podemos dar cuenta que al final se encuentra la matriz de varianzas y covarianzas de los $\beta's$.[10]

$$\widehat{Var(\hat{\beta})} = \frac{1}{(1415-8)} \begin{bmatrix} 4{,}7E-2 & 3{,}8E-3 & 9{,}7E-5 & 5{,}7E-1 & -4{,}7E-2 & -3{,}8E-3 & -9{,}7E-5 & -5{,}7E-1 \\ 3{,}8E-3 & 4{,}5E-2 & -8{,}6E-4 & 4{,}4E-1 & -3{,}8E-3 & -4{,}5E-2 & 8{,}6E-4 & -4{,}4E-1 \\ 9{,}7E-5 & -8{,}6E-4 & 1{,}9E-5 & -5{,}7E-3 & -9{,}7E-5 & 8{,}6E-4 & -1{,}9E-5 & 5{,}7E-3 \\ 5{,}7E-1 & 4{,}4E-1 & -5{,}7E-3 & 2{,}2E+1 & -1{,}0E+0 & -8{,}7E-1 & 1{,}1E-2 & -1{,}2E+1 \\ -4{,}7E-2 & -3{,}8E-3 & -9{,}7E-5 & -1{,}0E+0 & 9{,}0E-2 & 4{,}3E-3 & 1{,}9E-4 & 5{,}7E-1 \\ -3{,}8E-3 & -4{,}5E-2 & 8{,}6E-4 & -8{,}7E-1 & 4{,}3E-3 & 8{,}7E-2 & -1{,}6E-3 & 4{,}4E-1 \\ -9{,}7E-5 & 8{,}6E-4 & -1{,}9E-5 & 1{,}1E-2 & 1{,}9E-4 & -1{,}6E-3 & 3{,}4E-5 & -5{,}7E-3 \\ -5{,}7E-1 & -4{,}4E-1 & 5{,}7E-3 & -1{,}2E+1 & 5{,}7E-1 & 4{,}4E-1 & -5{,}7E-3 & 1{,}2E+1 \end{bmatrix}$$

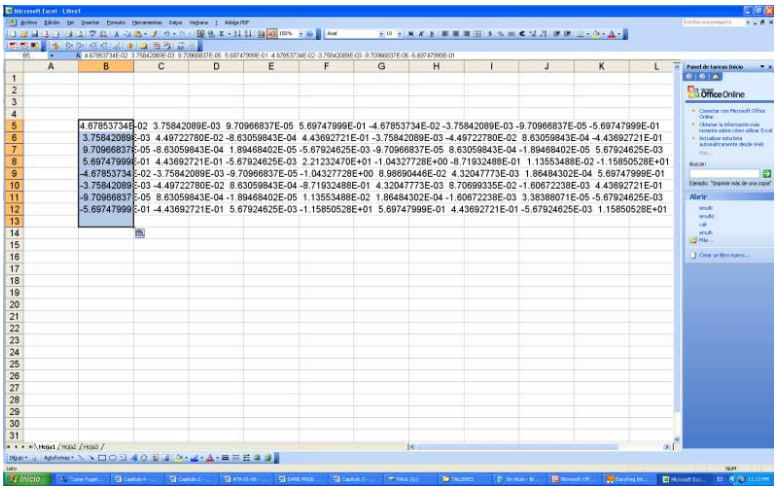
Con esta matriz podemos realizar la prueba de correlaciones de los betas. Para esto debemos copiar la matriz en Excel, sin embargo es mejor primero copiar en un bloc de notas, y de ser necesario se cambien puntos por comas o viceversa dependiendo de la configuración de su computador. Luego pegue los datos en Excel, observará que en la parte inferior aparece el asistente de pegado,[11] haga presione "usar el asistente para importar texto"

Es importante recordar que EasyReg reporta una matriz que no es exactamente la matriz de varianzas y covarianzas. Si se desea obtener la matriz de varianzas y covarianzas debemos dividir esta matriz por $n-k$.[12]

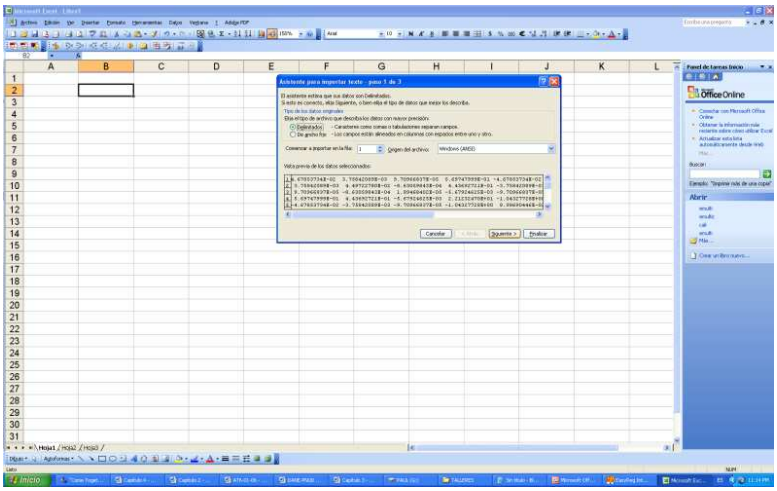
[10] Emplearemos la penúltima matriz que asume homoscedasticidad en los errores, la última de estas es útil en presencia de heteroscedasticidad, problema que trataremos en capítulos posteriores. La penúltima matriz corresponde a $SSE \times (X^TX)^{-1}$. Por lo tanto será necesario multiplicar por $\frac{1}{(n-k)}$ para obtener $\widehat{Var(\hat{\beta})}$. Además, noten que $n = 1415$ y $k = 8$

[11] Un icono en forma de maleta

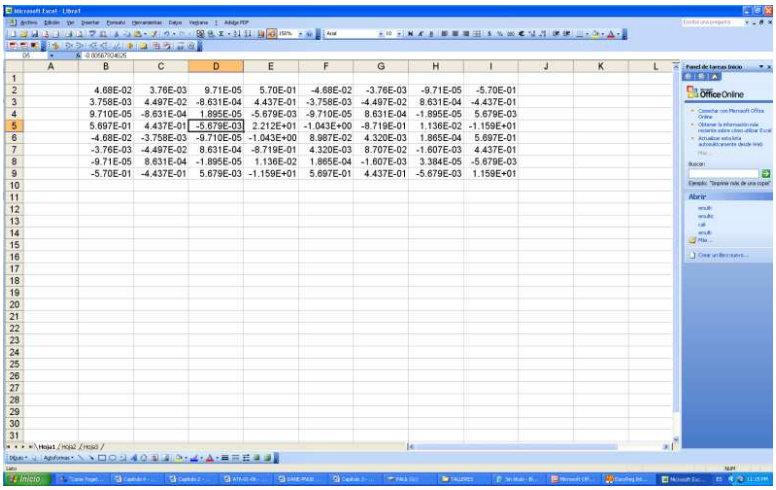
[12] Noten que en este caso se necesita dividir por n cada elemento de la penúltima matriz de EasyReg para obtener la matriz de varianzas y covarianzas. En este caso, eso no es necesario, pues al emplear la fórmula 4.5 se tiene: $\rho_{\hat{\beta}_1\hat{\beta}_2} = \frac{Cov(\hat{\beta}_1\hat{\beta}_2)}{\sqrt{Var(\beta_1)Var(\beta_2)}}$, entonces $\rho_{\hat{\beta}_1\hat{\beta}_2} = \frac{\frac{1}{n}(elemento\ (2{,}1))}{\sqrt{\frac{1}{n}elemento(1{,}1)\cdot\frac{1}{n}elemento(2{,}2)}} = \frac{\frac{1}{n}(elemento\ (2{,}1))}{\frac{1}{n}\sqrt{elemento(1{,}1)\cdot elemento(2{,}2)}}$



Escoja la opción delimitados y luego siguiente



En este caso los datos se encuentran separados por espacios, entonces conserve la opción que Excel tiene como predeterminada y presione nuevamente siguiente y finalizar. Como puede notar, ahora los datos han sido separados cada uno en celdas distintas, lo cual nos permite efectuar operaciones.



Empleando la fórmula 4.5 para calcular las correlaciones entre los coeficientes obtenemos las siguientes estimaciones:[13]

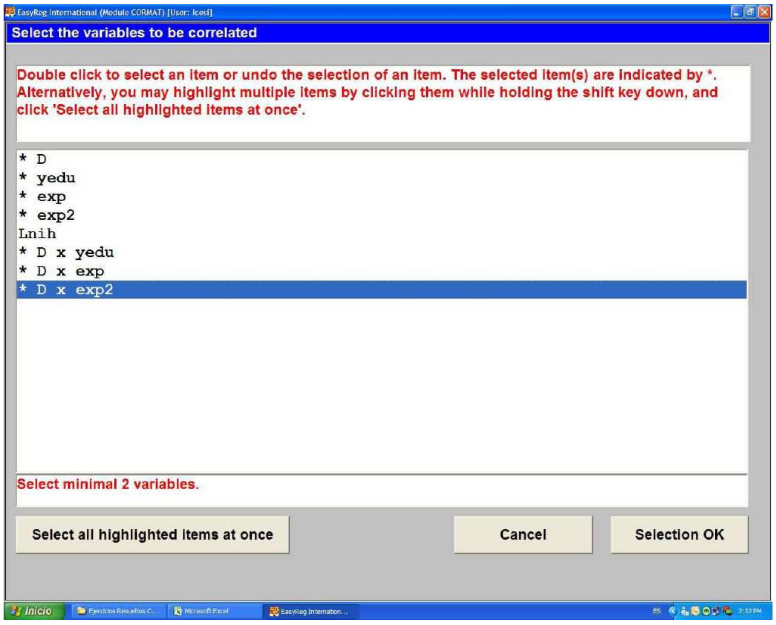
	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_0$
$\hat{\beta}_1$	1	0.08	0.10	0.56	-0.72	-0.06	-0.07	-0.77
$\hat{\beta}_2$	-	1	**-0.94**	0.44	-0.06	-0.71	0.70	-0.61
$\hat{\beta}_3$	-	-	1	-0.28	-0.07	0.67	-0.75	0.38
$\hat{\beta}_4$	-	-	-	1	-0.74	-0.63	0.42	-0.72
$\hat{\beta}_5$	-	-	-	-	1	0.04	0.11	0.56
$\hat{\beta}_6$	-	-	-	-	-	-	**-0.94**	0.44
$\hat{\beta}_7$	-	-	-	-	-	-	1	-0.29
$\hat{\beta}_0$	-	-	-	-	-	-	-	1

Como se aprecia la correlación entre los coeficientes asociados a las variables exp y exp^2 presentan una correlación en valor absoluto de 0.94 y entre los coeficientes asociados a $Dexp$ y $Dexp^2$ la correlación es de también de 0.94, ambas mayores de 0.8. Por lo tanto, esta prueba evidencia un problema de multicolinealidad grave en esta muestra entre estas dos variables.

Para realizar la prueba Kappa y la del determinante de la matriz de correlaciones de las variables explicativas, después de haber realizado la estimación presione "Done" lo cual lo llevará nuevamente al menú inicial. A continuación calcularemos la matriz de correlaciones de los $\hat{\beta}'s$, para esto elija "Menu" "Data analisis" y "Correlation matrix"

[13]Para que esta prueba tenga validez no debe existir heteroscedasticidad ni autocorrelación en la regresión. Como se verá en el siguiente capítulo esta regresión presenta un problema de heteroscedasticidad y por tanto esta prueba no tiene validez. Por ahora omitiremos este inconveniente

Luego haga doble clic todas las variables explicativas del modelo, EasyReg preguntará si desea realizar el análisis sobre una sub.-muestra, presione "No" y luego "Continue"



Luego EasyReg arrojará la estimación que se encuentra en el siguiente recuadro, de la cual vamos a necesitar los valores propios (Eigenvalues) de la matriz de correlaciones, que se encuentran al final de la estimación.

Recuadro 4.3 Matriz de correlaciones

Variables:
X(1)=D
X(2)=yedu
X(3)=exp
X(4)=exp2
X(5)=D x yedu
X(6)=D x exp
X(7)=D x exp2

First chosen observation: t = 1
Last chosen observation: t = 1415

Variable	Sample mean	Sample standard error
X(1)	0.4911661	0.5000987
X(2)	9.6261484	4.2875129
X(3)	19.1003534	12.4080707
X(4)	518.6749117	634.8057274
X(5)	4.7328622	5.6595429
X(6)	9.5399293	13.0341596
X(7)	260.7795053	535.4391654

Sample second moments matrix (X'X/n)

4.91E-01	4.73E+00	9.54E+00	2.60E+02	4.739E+00	9.54E+00	2.61E+02
4.73E+00	1.11E+02	1.61E+02	3.81E+03	5.44E+01	8.30E+01	2.01E+03
9.54E+00	1.61E+02	5.19E+02	1.74E+04	8.30E+01	2.61E+02	8.89E+03
2.61E+02	3.82E+03	1.74E+04	6.72E+05	2.01E+03	8.89E+03	3.54E+05
4.73E+00	5.44E+01	8.30E+01	2.01E+03	5.44E+01	8.30E+01	2.01E+03
9.54E+00	8.30E+01	2.61E+02	8.89E+03	8.30E+01	2.61E+02	8.89E+03
2.61E+02	2.01E+03	8.89E+03	3.54E+05	2.01E+03	8.89E+03	3.54E+05

Eigenvalues:

9.02E+05	1.25E+05	1.89E+02	2.78E+01	1.20E+01	3.00E+00	3.12E-02

Orthogonal matrix of eigenvectors:

4.00E-04	6.17E-04	3.24E-02	4.84E-02	-1.63E-02	2.22E-03	9.98E-01
4.77E-03	-3.08E-03	6.37E-01	-5.75E-01	-3.98E-01	3.24E-01	1.02E-05
2.15E-02	-1.59E-02	5.54E-01	-1.12E-01	7.12E-01	-4.15E-01	3.09E-05
8.39E-01	-5.44E-01	-1.84E-02	6.25E-03	-1.68E-02	9.25E-03	-8.84E-07
3.09E-03	4.77E-03	3.93E-01	5.05E-01	-5.42E-01	-5.42E-01	-4.49E-02
1.36E-02	2.10E-02	3.62E-01	6.32E-01	1.99E-01	6.54E-01	-4.06E-02
5.44E-01	8.38E-01	-1.04E-02	-1.90E-02	-6.97E-04	-1.40E-02	5.39E-04

Standardized eigenvectors:

4.77E-04	7.36E-04	5.08E-02	7.66E-02	-2.29E-02	3.39E-03	1.00E+00
5.68E-03	-3.68E-03	1.00E+00	-9.10E-01	-5.58E-01	4.95E-01	1.02E-05
2.57E-02	-1.90E-02	8.69E-01	-1.77E-01	1.00E+00	-6.34E-01	3.09E-05
1.00E+00	-6.48E-01	-2.89E-02	9.90E-03	-2.36E-02	1.42E-02	-8.85E-07
3.69E-03	5.69E-03	6.16E-01	7.98E-01	-7.61E-01	-8.30E-01	-4.50E-02
1.63E-02	2.51E-02	5.68E-01	1.00E+00	2.79E-01	1.00E+00	-4.07E-02
6.48E-01	1.00E+00	-1.63E-02	-2.99E-02	-9.79E-04	-2.14E-02	5.40E-04

Sample variance matrix

22.50E-01	4.83E-03	1.59E-01	6.03E+00	2.41E+00	4.86E+00	1.33E+02
4.83E-03	1.84E+01	-2.28E+01	-1.18E+03	8.85E+00	-8.83E+00	-4.96E+02
1.59E-01	-2.28E+01	1.54E+02	7.47E+03	-7.39E+00	7.86E+01	3.91E+03
6.03E+00	-1.18E+03	7.47E+03	4.03E+05	-4.40E+02	3.94E+03	2.19E+05
2.41E+00	8.85E+00	-7.39E+00	-4.40E+02	3.20E+01	3.79E+01	7.81E+02
4.86E+00	-8.83E+00	7.86E+01	3.94E+03	3.79E+01	1.69E+02	6.40E+03
1.33E+02	-4.96E+02	3.91E+03	2.19E+05	7.81E+02	6.40E+03	2.87E+05

Eigenvalues:

5.72E+05 1.18E+05 4.32E+01 2.05E+01 1.19E+01 2.99E+00 1.48E-02

Orthogonal matrix of eigenvectors:

1.50E-04	8.62E-04	4.98E-02	2.07E-03	-4.84E-02	4.21E-04	9.98E-01
-2.16E-03	2.76E-03	1.29E-01	6.30E-01	6.87E-01	-3.36E-01	2.57E-02
1.45E-02	-1.23E-02	2.67E-01	-5.66E-01	6.66E-01	4.05E-01	2.00E-02
7.92E-01	-6.10E-01	1.10E-03	1.34E-02	-1.2E-02	-9.10E-03	-2.57E-04
2.23E-04	7.53E-03	6.82E-01	4.20E-01	-2.46E-01	5.44E-01	-4.71E-02
1.23E-02	2.27E-02	6.66E-01	-3.25E-01	-1.45E-01	-6.53E-01	-3.93E-02
6.10E-01	7.92E-01	-2.10E-02	4.64E-03	5.25E-03	1.40E-02	5.13E-04

Standardized eigenvectors:

1.89E-04	1.09E-03	7.29E-02	3.29E-03	-7.05E-02	6.44E-04	1.00E+00
-2.73E-03	3.48E-03	1.90E-01	1.00E+00	1.00E+00	-5.14E-01	2.58E-02
1.83E-02	-1.56E-02	3.91E-01	-8.97E-01	9.70E-01	6.19E-01	2.01E-02
1.00E+00	-7.70E-01	1.62E-03	2.13E-02	-1.75E-02	-1.39E-02	-2.57E-04
2.81E-04	9.50E-03	1.00E+00	6.65E-01	-3.58E-01	8.32E-01	-4.72E-02
1.55E-02	2.86E-02	9.76E-01	-5.16E-01	-2.11E-01	-1.00E+00	-3.94E-02
7.70E-01	1.00E+00	-3.08E-02	7.36E-03	7.65E-03	2.14E-02	5.14E-04

Sample correlation matrix

1.00E+00	2.25E-03	2.56E-02	1.90E-02	8.51E-01	7.45E-01	4.96E-01
2.25E-03	1.00E+00	-4.28E-01	-4.33E-01	3.65E-01	-1.58E-01	-2.16E-01
2.56E-02	-4.27E-01	1.00E+00	9.48E-01	-1.05E-01	4.86E-01	5.89E-01
1.90E-02	-4.33E-01	9.48E-01	1.00E+00	-1.23E-01	4.77E-01	6.45E-01
8.51E-01	3.65E-01	-1.05E-01	-1.23E-01	1.00E+00	5.14E-01	2.58E-01
7.45E-01	-1.58E-01	4.86E-01	4.77E-01	5.14E-01	1.00E+00	9.18E-01
4.96E-01	-2.16E-01	5.89E-01	6.45E-01	2.578E-01	9.18E-01	1.00E+00

Eigenvalues:

3.45E+00 2.28E+00 7.28E-01 4.08E-01 7.18E-02 4.45E-02 9.48E-03

Orthogonal matrix of eigenvectors:

3.29E-01	4.75E-01	-3.26E-01	-1.77E-01	1.10E-01	6.79E-01	-2.36E-01
-1.86E-01	3.84E-01	8.56E-01	1.15E-01	1.13E-01	2.41E-01	-8.93E-03
4.01E-01	-3.65E-01	2.66E-01	-3.99E-01	5.55E-01	-1.39E-01	-3.85E-01
4.06E-01	-3.72E-01	2.72E-01	-2.89E-01	-5.18E-01	3.39E-01	3.94E-01
2.05E-01	5.66E-01	5.05E-02	-5.02E-01	-2.85E-01	-5.49E-01	1.85E-02
4.97E-01	1.85E-01	-3.12E-02	3.48E-01	4.27E-01	-1.45E-01	6.27E-01
4.93E-01	1.02E-02	1.11E-01	5.82E-01	-3.67E-01	-1.54E-01	-4.97E-01

Standardized eigenvectors:

6.61E-01	8.39E-01	-3.80E-01	-3.05E-01	1.97E-01	1.00E+00	-3.77E-01
-3.74E-01	6.79E-01	1.00E+00	1.98E-01	2.04E-01	3.55E-01	-1.42E-02
8.06E-01	-6.45E-01	3.11E-01	-6.86E-01	1.00E+00	-2.04E-01	-6.15E-01
8.17E-01	-6.58E-01	3.17E-01	-4.97E-01	-9.33E-01	4.99E-01	6.29E-01
4.12E-01	1.00E+00	5.89E-02	-8.63E-01	-5.13E-01	-8.09E-01	2.95E-02
1.00E+00	3.28E-01	-3.65E-02	5.99E-01	7.70E-01	-2.14E-01	1.00E+00
9.92E-01	1.80E-02	1.29E-01	1.00E+00	-6.62E-01	-2.27E-01	-7.93E-01

Remark: Numbers with exponent E-15 or more negative should be interpreted as zero.

Para realizar la prueba Kappa, en este caso tenemos que los valores propios (eigenvalues) máximos y mínimos de la matriz son: $\lambda_1 = 3{,}45433$ y $\lambda_7 = 0{,}00948$ por tanto:

$$\kappa\left(X\right) = \frac{\sqrt{\lambda_{MAX}}}{\sqrt{\lambda_{MIN}}} = \sqrt{\frac{3{,}45443}{0{,}00948}} = 19{,}09 \tag{4.7}$$

Como podemos apreciar en la ecuación 4.7 el valor de $k(X)$ no tiende a 1 pero no es mayor a 20. Así existe un problema de Multicolinealidad no muy grave.

Finalmente calculamos el determinante de la matriz de correlación $|\mathbf{RX}|$ empleando los valores propios, con lo cual se obtiene que:

$$|\mathbf{RX}| = 3{,}4544 \times 2{,}2838 \times 0{,}7281 \times 0{,}4079 \times 0{,}0718 \times 0{,}0445 \times 0{,}0095 = 0{,}0000708 \tag{4.8}$$

Como este valor tiende a cero, concluimos de acuerdo a esta prueba evidencia la presencia de multicolinealidad grave.

Finalmente y teniendo en cuenta los resultados de todas las pruebas efectuadas, podemos concluir que el modelo presenta una alta correlación entre las variables exp y exp^2. De la definición de las variables sabemos que no existe una relación lineal perfecta, pero nuestro hallazgo puede ser síntoma de que los valores de la variable exp corresponden a un rango relativamente corto, para el cual la relación exponencial puede ser aproximada por una relación lineal. Ahora bien, no estamos frente a un problema de multicolinealidad perfecta, y por lo tanto no se está violando el segundo supuesto del teorema de Gauss Markov. Por otro lado, el modelo teórico necesita incluir en la especificación ambas variables por razones bien fundamentadas en Oaxaca (1973) y sería incorrecto eliminar alguna de las dos variables. En el siguiente capítulo continuaremos con el análisis de esta regresión.

4.5. Ejercicios

Un investigador es contratado para estimar un modelo sencillo que explique, empleando la teoría económica, la tasa de crecimiento de las importaciones de un pequeño país del Pacífico Sur. Para elaborar la tarea se cuenta con 27 observaciones (la información se encuentra en el archivo multi.xls). Las observaciones corresponden a las series de Estados Unidos desde 1990 hasta el 2004 de las tasas de crecimiento de las importaciones (Y_t), el producto interno bruto (X_{1t}) y la inflación(X_{2t}). Al investigador que de inmediato se le asigna esta tarea, realiza una exhaustiva revisión bibliográfica que le lleva a plantear el siguiente modelo para explicar el comportamiento del crecimiento de las exportaciones:

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 X_{2t} + \varepsilon_t \tag{4.9}$$

1. Estime el modelo 4.9 y explique si existen problemas de multicolinealidad
2. Determine si existe o no multicolinealidad por medio de las pruebas estudiadas en este capítulo
3. Estime dos regresiones simples a partir del modelo 4.9; es decir, la tasa de crecimiento de las importaciones en función de sólo una de las variables independientes mediante los siguientes modelos:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \varepsilon_i \tag{4.10}$$

$$Y_i = \beta_0 + \beta_1 X_{2i} + \varepsilon_i \tag{4.11}$$

¿Qué ocurre con la significancia y el ajuste del modelo?

4. Teniendo en cuenta los resultados hasta ahora encontrados, corrija (de ser posible) el problema de multicolinealidad si es que existe. Explique.

CAPÍTULO 5

Heteroscedasticidad

Objetivos del capítulo

Al finalizar este capítulo, el lector estará en capacidad de:

- Realizar diferentes tipos de análisis gráficos que revelen la posibilidad de heteroscedasticidad en los residuos empleando EasyReg.
- Efectuar mediante EasyReg tres pruebas estadísticas necesarias para detectar la violación del supuesto de homoscedasticidad en los residuos. En especial las pruebas de Goldfeld y Quant, Breusch-Pagan y la prueba de White.
- Estimar con EasyReg la matriz de varianzas y covarianzas consistente en presencia de heteroscedasticidad de White.

5.1. Introducción

Como lo hemos discutido en capítulos anteriores, si los supuestos del modelo de regresión que se resumen en el recuadro 5.1 se cumplen, entonces el Teorema de Gauss-Markov demuestra que los estimadores MCO son MELI (Mejor Estimador Lineal Insesgado). En el Capítulo 4 analizamos las consecuencias de la violación del supuesto de independencia lineal entre variables explicativas en un modelo de regresión múltiple. En este Capítulo nos concentramos en la violación del supuesto que el término de error tiene varianza constante.

Recuadro 5.1 Teorema de Gauss-Markov

Si se considera un modelo lineal $\mathbf{y_{n\times 1}} = \mathbf{X_{n\times k}}\beta_{k\times 1} + \varepsilon_{n\times 1}$ y se supone que:

- Las $X_2, X_3, ..., X_k$ son fijas y linealmente independientes (es decir X tiene rango completo y es una matriz no estocástica).
- El vector de errores ε tiene media cero, varianza constante y no autocorrelación. Es decir: $E[\varepsilon] = 0$ y $Var[\varepsilon] = \sigma^2 I_n$

Entonces el estimador de MCO $\hat{\beta} = \left(X^T X\right)^{-1} X^T \mathbf{y}$ es el *Mejor Estimador Lineal Insesgado (MELI)*

En ocasiones el supuesto de homoscedasticidad o varianza del término de error constante no tiene mucho sentido. Un ejemplo de esto se presenta al analizar el consumo en función del ingreso con una muestra de hogares (corte transversal). Supongamos que se desea emplear un modelo en que el consumo del hogar i depende de su nivel de ingresos del respectivo hogar. Los hogares con ingresos bajos presenten un comportamiento en su consumo mucho menos variable que el consumo de los hogares con altos ingresos. Es decir, es de esperarse que para ingresos altos el comportamiento del consumo se disperse más con respecto a lo esperado (su media). Por otro lado, a medida que el ingreso sea más bajo los hogares no podrán tener una dispersión muy grande con respecto a lo esperado para su nivel de ingresos. Esto implica que las observaciones correspondientes al consumo de hogares con ingresos bajos tendrán una varianza con respecto a su valor esperado (varianza del error) mucho menor que aquellos hogares con ingresos altos. También existen otras razones para que se presente la heteroscedasticidad, como por ejemplo el aprendizaje sobre los errores y mejoras en la recolección de la información a medida que ésta se realiza.

En general, este problema econométrico es muy común en datos de corte transversal, aunque es posible que el problema también se presente con series de tiempo. Formalmente, en presencia de este problema, la matriz de varianzas y covarianzas de los errores será:

$$Var[\varepsilon] = E\left[\varepsilon^T \varepsilon\right] = \Omega = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \neq \sigma^2 I$$

En presencia de heteroscedasticidad, los estimadores MCO siguen siendo insesgados, pero ya no tienen

la mínima varianza posible.[1] Es decir los estimadores MCO no son MELI.

De hecho, el estimador MCO de la matriz de varianzas y covarianzas ($Var\left[\hat{\beta}\right] = \sigma^2\left(\mathbf{X^T X}\right)^{-1}$) será sesgado. En presencia de heteroscedasticidad la matriz de varianzas y covarianzas del estimador MCO del vector β es:[2]

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\Omega\mathbf{X}\left(\mathbf{X^T X}\right)^{-1}$$

Esto implica que el estimador de la matriz de varianzas y covarianzas $\widehat{Var\left[\hat{\beta}\right]} = s^2\left(\mathbf{X^T X}\right)^{-1}$ sea sesgado y por lo tanto si usamos este último estimador en pruebas de hipótesis o intervalos de confianza para los coeficientes estimados, entonces obtendremos conclusiones erróneas en torno a los verdaderos $\beta's$. Esto ocurrirá en las pruebas individuales y conjuntas.

Así, si se emplea el estimador MCO en presencia de heteroscedasticidad, entonces los estimadores de los coeficientes serán insesgados, pero los errores estándar no serán los adecuados. Por tanto, cualquier conclusión derivada de la inferencia a partir de los estimadores MCIO será incorrecta.

5.2. Pruebas para la detección de heteroscedasticidad

En general, una buena práctica cuando se calculan modelos econométricos es emplear gráficos que permitan intuir que está ocurriendo con los residuos estimados. Obviamente, los gráficos no proveen evidencia contundente para concluir, pero si provee la intuición necesaria para iniciar las pruebas formales.

Dado que la heteroscedasticidad implica una variabilidad no constante del error, entonces lo más adecuado sería poder graficar el vector de errores (ε). Lastimosamente, este vector no es observable, pero la mejor aproximación que tenemos para conocer ese vector sería el error estimado ($\hat{\varepsilon}$).

Así, el primer paso para detectar la presencia de heteroscedasticidad es graficar los residuos del modelo estimado. Los gráficos más empleados son gráficos de dispersión de:

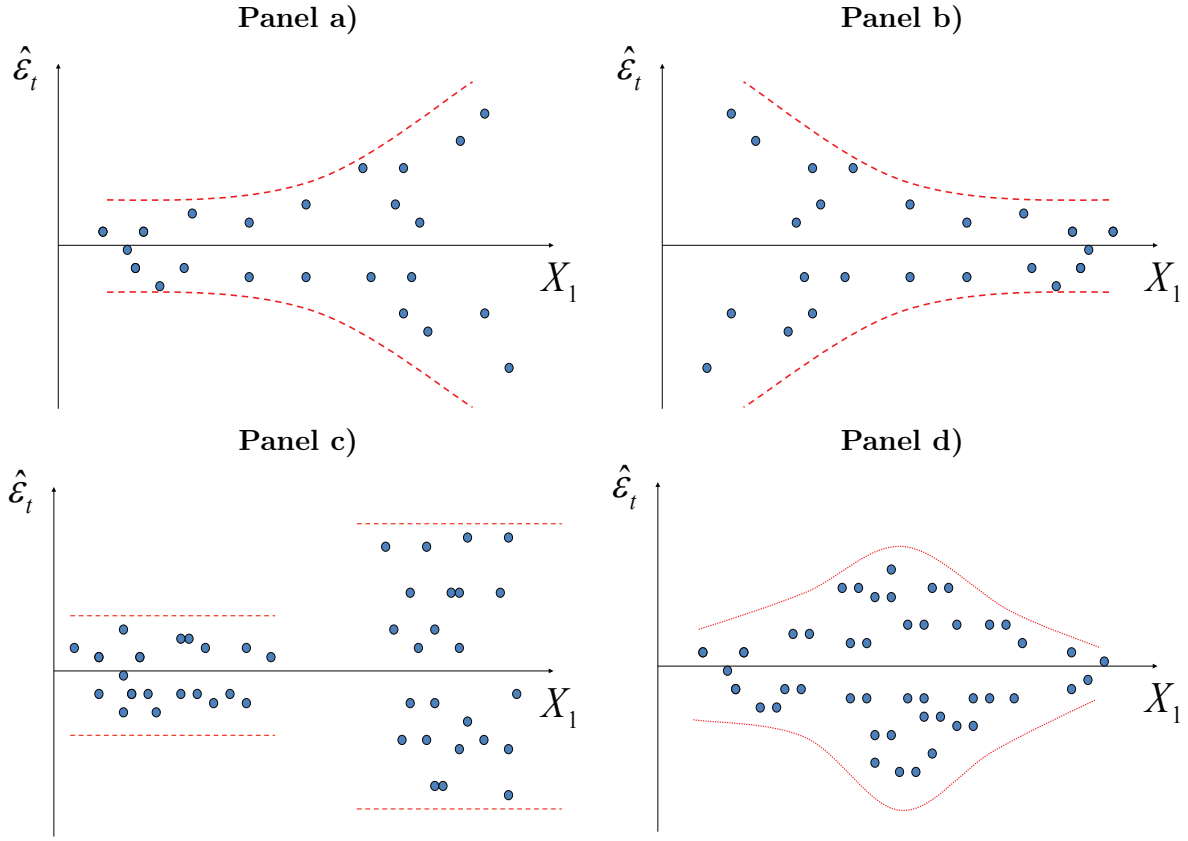
1. Los errores estimados versus las observaciones ($\hat{\varepsilon}_i$ vs $i = 1, 2, ...n$)

2. Los errores estimados versus cada una de las variables explicativas ($\hat{\varepsilon}_i$ vs $\ bfX's$)

En estos gráficos se busca algún tipo de regularidad como la que se presenta en la figura 5.1. En el panel a) observamos que los residuos se vuelven más grandes a medida que la variable dependiente crece. En este caso la dispersión con respecto a la media (cero) crece a medida que la variable explicativa crece y esto sucede de manera exponencial. Por tanto, existe heteroscedasticidad que podría ser del tipo $\sigma_i^2 = X_{1i}\sigma^2$. En el panel b) el comportamiento de la varianza es opuesto al presentado en el panel a), heteroscedaticidad que podría ser del tipo $\sigma_i^2 = \frac{1}{X_{1i}}\sigma^2$. En el panel c) claramente existen dos grupos de datos; los de varianza grande y los de baja varianza. Por último en el panel d) los residuos son menores para las mediciones grandes y pequeñas pero crecen en los valores intermedios de la misma. En los cuatro casos hay síntomas de heteroscedasticidad.

[1] En el Apéndice 5.1 se presenta una demostración.
[2] En el Apéndice 5.2 se presenta una demostración.

Figura 5.1: Tipos de heteroscedasticidad



Sin embargo, el análisis gráfico es sólo análisis informal que nos permite intuir la presencia del problema. Una detección formal de la heteroscedasticidad implica emplear pruebas estadísticas. A continuación consideraremos tres pruebas, cada una diseñada para detectar diferentes "tipos"de heteroscedasticidad; partiendo de los casos más particulares al más general de heteroscedasticidad. En todos los casos la hipótesis nula de estas pruebas es la presencia de homoscedasticidad $H_0 : \sigma_i^2 = \sigma^2; \forall i$ versus la hipótesis alterna de que existe algún tipo de heteroscedasticidad.

5.2.1. Prueba de Goldfeld y Quandt

Goldfeld y Quandt (1965) diseñaron una prueba de heteroscedasticidad que supone una relación entre la varianza de cada error y una variable del tipo $\sigma_i^2 = \sigma^2 X_i^2$. De tal manera que esta prueba implica las siguientes hipótesis nula y alterna:

$$\begin{aligned} H_0 &: \sigma_i^2 = \sigma^2; \ \ \forall i \\ H_A &: \sigma_i^2 = \sigma^2 X_{j,i}^2 \end{aligned} \tag{5.1}$$

Para comprobar esta hipótesis, los autores proponen dividir las observaciones en dos grupos, los de alta (grupo 1) y baja varianza (grupo 2). De tal forma que si la hipótesis nula es correcta entonces las varianzas de ambos grupos deberán ser las mismas. Es decir si se cumple que $Var\left[\hat{\varepsilon}_1\right] \approx Var\left[\hat{\varepsilon}_2\right]$ entonces no podremos rechazar la hipótesis nula. Con esta idea en mente Goldfeld y Quandt (1965) proponen

dividir la suma cuadrada de los errores de ambos grupos. Ellos proponen el siguiente estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1}$$

Finalmente, bajo el supuesto de que los errores siguen una distribución normal, Goldfeld y Quandt (1965) demostraron que el estadístico F_{GQ} sigue una distribución $F_{(n-d-2k,n-d-2k)}$. Por tanto se rechazará la hipótesis nula de homoscedasticidad si $F_{GQ} > F_{(n-d-2k,n-d-2k),\alpha}$ con un nivel de confianza del $(1-\alpha)\,\%$.

Los pasos para efectuar esta prueba son los siguientes:

1. Ordenar los datos de menor a mayor de acuerdo a la variable X_j (que se sospecha que está relacionada con la varianza del error)[3]
2. Omita los d datos de la mitad teniendo en cuenta que $d < \frac{1}{5}n$
3. Corra dos regresiones: una para el grupo de datos de baja varianza ($R1$) y una para el grupo de datos de alta varianza ($R2$)
4. Encuentre la suma cuadrada de los residuos de cada una de las regresiones (SSE_1 y el SSE_2)
5. Calcule el estadístico de Goldfeld y Quandt, de la siguiente manera[4] $F_{GQ} = \frac{SSE_2}{SSE_1}$
6. Compare el estadístico F_{GQ} con $F_{(n-d-2k,n-d-2k)}$.

La intención de eliminar los d datos de la mitad es acentuar la diferencia entre las varianzas de los dos grupos. No eliminar los datos de la mitad puede afectar el poder de la muestra. No obstante, en algunos casos puede no ser necesario eliminar datos, como por ejemplo el caso presentado en el panel c) de la figura 5.1

Antes de discutir la siguiente prueba, es importante destacar que esta prueba funciona bien si es posible ordenar los datos, como por ejemplo en los casos presentados en los paneles a), b) y c) de la figura 5.1. En esos casos es posible dividir la muestra en dos y apreciar cambios significativos en la varianza del error.

Sin embargo en el caso que se muestra en el panel d) no sería muy útil esta prueba, ya que al eliminar los datos de la mitad, los datos de los grupos restantes tendrán prácticamente la misma varianza. Así, esta prueba tiene poder para detectar la presencia de heteroscedasticidad del tipo $\sigma_i^2 = \sigma^2 X_i^2$.

5.2.2. Prueba de Breusch-Pagan

Esta prueba, a diferencia del test de Goldfeld y Quandt, no requiere que las observaciones sean ordenadas en función de la variable independiente que consideramos que pueda estar relacionada con la varianza del error y permite la posibilidad de que más de una variable cause el problema de heteroscedasticidad.

Breusch y Pagan (1979) diseñaron una prueba que permite detectar sí existe una relación entre la varianza del error y un grupo de variables (recogidas en un vector Z). El tipo de relación de esta prueba

[3]Esta variable se puede identificar empleando los gráficos. En casos en que los gráficos no permitan determinar claramente cuál es la variable que causa el problema, será mejor realizar la prueba con diferentes variables.

[4]Note que en todos los casos tendremos que $SSR_1 < SSR_2$.

no está suscrita a algún tipo de relación funcional como si ocurre en la prueba de Goldfeld y Quandt. En este caso, la prueba está diseñada para detectar la heteroscedasticidad de la forma $\sigma_i^2 = f(\gamma + \delta Z_i)$. Donde $Z_{i(g\times 1)}$ es un grupo de g variables que afectan a la varianza que se organizan en forma vectorial y $\delta_{(1\times g)}$ corresponde a un vector de constantes.

Por ejemplo, supongamos que el investigador cree que el problema de heteroscedasticidad está siendo causado por las variables W_i y V_i. En ese caso tendremos que $Z_{i(2\times 1)} = [W_i, V_i]^T$ y $\delta_{(1\times 2)} = [\delta_1, \delta_2]$. Por tanto, $\sigma_i^2 = f(\gamma + \delta_1 W_i + \delta_2 V_i)$.

Recapitulando, la prueba de Breusch-Pagan implica las siguientes hipótesis nula y alterna:

$$H_0 : \sigma_i^2 = \sigma^2; \ \forall i$$
$$H_A : \sigma_i^2 = f(\gamma + \delta Z_i)$$

Los pasos para efectuar esta prueba son los siguientes:

1. Corra el modelo original $Y = X\beta + \varepsilon$ y encuentre la serie de los residuos $\hat{\varepsilon}$.
2. Calcule[5] $\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n} = \frac{\sum_{i=1}^{n}\hat{\varepsilon}_i^2}{n}$.
3. Posteriormente estime la siguiente regresión auxiliar: $\frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta Z_i + \mu_i$.
4. Calcule la suma de los cuadrados de la regresión auxiliar ($SSR = SST - SSE$).
5. Calcule el estadístico $BP = \frac{SSR}{2}$.

Breusch y Pagan (1979) demostraron que el estadístico BP sigue una distribución Chi-cuadrado con g grados de libertad (χ_g^2), bajo el supuesto de que los errores se distribuyen normalmente. Por tanto, se rechazará la hipótesis nula de homoscedasticidad a favor de una heteroscedasticidad de la forma $\sigma_i^2 = f(\gamma + \delta Z_i)$ si el estadístico BP es mayor que $\chi_{g,\alpha}^2$, con un nivel de confianza del $(1-\alpha)\,\%$.

5.2.3. Prueba de White

White (1980) desarrolló una prueba más general que las anteriores, con la ventaja de no requerir que ordenemos los datos en diferentes grupos, ni tampoco depende de que los errores se distribuyan normalmente. Esta prueba implica las siguientes hipótesis nula y alterna:

$$H_0 : \sigma_i^2 = \sigma^2; \ \forall i$$
$$H_A : No\ H_0$$

Los pasos para efectuar esta prueba son los siguientes:

1. Corra el modelo original $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ y encuentre la serie de los residuos $\hat{\varepsilon}$.

[5]Noten que esto corresponde al estimador de la varianza del error del Método de Máxima Verosimilitud.

2. Ahora corra la siguiente regresión auxiliar: $\hat{\varepsilon}_i^2 = \gamma + \sum_{m=1}^{k}\sum_{j=1}^{k}\delta_s X_{mi}X_{ji} + \mu_i$. Por ejemplo, si el modelo original es $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$, entonces el modelo para la regresión auxiliar sería:[6] $\hat{\varepsilon}_i^2 = \gamma + \delta_1 X_{2i} + \delta_2 X_{3i} + \delta_3 X_{2i}^2 + \delta_4 X_{3i}^2 + \delta_5 X_{2i}X_{3i} + \vartheta_i$.
3. Calcule el R^2 de la regresión auxiliar.
4. Calcule el estadístico de White $W_a = n \times R^2$.

White (1980) demostró que su estadístico W_a sigue una distribución Chi-cuadrado con un número de grados de libertad igual al número de regresores que se emplean en la regresión auxiliar (g). Por tanto se rechazará la hipótesis nula de no heteroscedasticidad con un nivel de confianza de $(1-\alpha)\,\%$, cuando el estadístico de esta prueba sea mayor que $\chi^2_{g,\alpha}$.

5.3. Solución a la heteroscedasticidad

En la sección pasada se discutió cómo detectar la presencia de heteroscedasticidad; pero, ¿qué hacer si ésta está presente en un modelo de regresión? A continuación, se discuten dos soluciones muy empleadas en la actualidad.

5.3.1. Solución por mínimos cuadrados ponderados

Si conocemos la varianza de cada uno de los errores es posible utilizar el método de Mínimos Cuadrados Ponderados (MCP) que hace parte de la familia de los Mínimos Cuadrados Generalizados (MCG). La idea de esta solución es muy sencilla, e implica transformar los datos de la muestra de tal forma que el problema desaparezca de la muestra. Por ejemplo, partamos del siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \varepsilon_i \tag{5.4}$$

Con un vector de errores con las siguientes características:$E[\varepsilon_i] = 0$, $E[\varepsilon_i \varepsilon_j] = 0$, y con una varianza de los errores conocida $Var[\varepsilon_i] = W_i^2\sigma_i^2$. Si dividimos el modelo por W_i obtenemos:[7]

$$\frac{Y_i}{W_i} = \beta_1\frac{1}{W_i} + \beta_2\frac{X_{2i}}{W_i} + \beta_3\frac{X_{3i}}{W_i} + \ldots + \beta_k\frac{X_{ki}}{W_i} + \frac{\varepsilon_i}{W_i} \tag{5.5}$$

Note que:

$$Var\left[\frac{\varepsilon_i}{W_i}\right] = \frac{1}{W_i^2}Var[\varepsilon_i] = \frac{1}{W_i^2}W_i^2\sigma^2 = \sigma^2$$

Es decir, ahora cada observación tiene varianza constante y por lo tanto el problema de heteroscedasticidad ha sido resuelto, de tal manera que los estimadores MCO para el modelo 5.5 ya son MELI.

El problema práctico de esta aproximación es conocer exactamente la naturaleza de la heteroscedasticidad. Ese problema puede ser resuelto por medio de la prueba de Goldfeld y Quandt. Ésta es la única de las tres vistas anteriormente que permite determinar con exactitud el tipo de heteroscedasticidad, de tal forma que el método de MCP se convierta en viable.

[6] El último término se conoce con el nombre de término cruzado. En algunas oportunidades cuando el modelo original cuenta con muchas variables explicatorias y/o el número de observaciones no es mucho, puede ocurrir que no existan los grados de libertad necesarios para correr la regresión auxiliar incluyendo los términos cruzados. En esos casos, algunos autores acostumbran correr la regresión auxiliar sin los términos cruzados, si bien esto le resta poder a la prueba.

[7] Los W_i son conocidos como los pesos de ponderación, de ahí el nombre del método.

5.3.2. Corrección de White

Es muy probable que en la práctica no podamos encontrar la forma exacta de la heteroscedasticidad y por tanto no podremos "corregir" la muestra, de tal manera que la heteroscedasticidad desaparezca, tal como lo hace el método de MCP. Para dar solución a esta dificultad, White (1980) ideó una forma de corregir el estimador y no la muestra.

En especial, White (1980) mostró que es posible aún encontrar un estimador apropiado para la matriz de varianzas y covarianzas de los $\beta's$ obtenidos por MCO.[8] Recordemos que en presencia de heteroscedasticidad, la varianza de los coeficientes tiene la siguiente estructura:

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\Omega\mathbf{X}\left(\mathbf{X^T X}\right)^{-1}$$

White (1980) sugiere usar el siguiente estimador consistente para la matriz de varianzas y covarianzas de los $\beta's$ obtenidos por MCO

$$Est.Var\left[\hat{\beta}\right] = n\left(\mathbf{X^T X}\right)^{-1}S_0\left(\mathbf{X^T X}\right)^{-1}$$

donde:

$$S_0 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 x_i x_i^T \qquad x_i^T = \begin{pmatrix} 1 & x_{1i} & x_{2i} & \dots & x_{ki} \end{pmatrix}$$

Esa matriz de varianzas y covarianzas deberá ser empleada para calcular los estadísticos de las pruebas de hipótesis tanto individuales como conjuntas. De esta manera, el problema del sesgo en la matriz de varianzas y covarianzas es solucionado. De hecho, White (1980) demostró que ese estimador es consistente.

5.4. Análisis del efecto discriminatorio de género en las diferencias salariales en Colombia

El objetivo de este ejercicio es aplicar las diferentes pruebas de heteroscedasticidad y de ser el caso aplicar la solución más conveniente a este problema. Continuaremos con el análisis del efecto discriminatorio de género en las diferencias salariales en Colombia. En el Capítulo 5 ya habíamos realizado una aproximación teórica a este modelo y aplicado las distintas pruebas de multicolinealidad. El modelo a estimar es el siguiente:[9]

$$\begin{aligned}\ln(ih_i) =& \beta_0 + \beta_1 yedu_i + \beta_2 exp_i + \beta_3 exp_i^2 + \beta_4 D_i \\ &+ \beta_5 Dyedu_i + \beta_6 Dexp_i + \beta_7 Dexp_i^2 + \varepsilon_i\end{aligned} \tag{5.6}$$

donde

$$D_t = \begin{cases} 1 & \text{si el individuo } i \text{ es hombre} \\ 0 & o.w. \end{cases}$$

$\ln(ih_i)$ representa el logaritmo natural del ingreso por hora del individuo i, $yedu_i$ y exp_i denotan los años de educación y de experiencia del individuo i. Los datos se encuentran en el archivo emulti.xls.

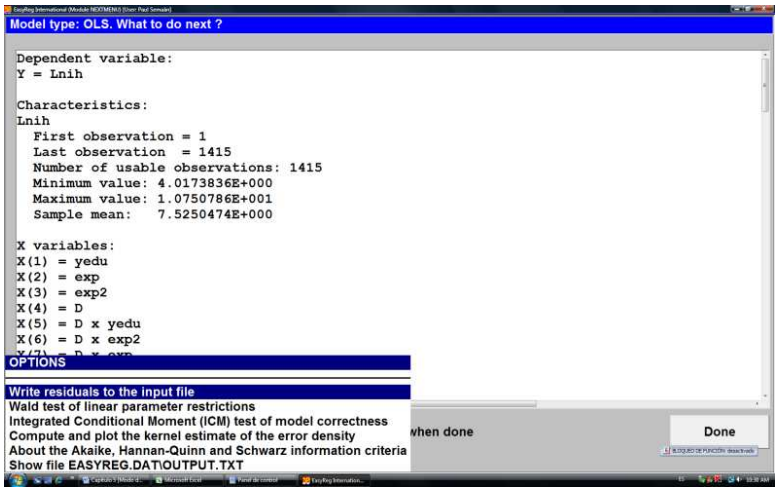
[8] Recuerden que el estimador MCO del vector β sigue siendo insesgado, el problema se presenta en el estimador de la matriz de varianzas y covarianzas.

[9] Si desea información teórica acerca de este modelo, remítase al Capítulo 4.

5.4.1. Análisis gráfico de los residuos

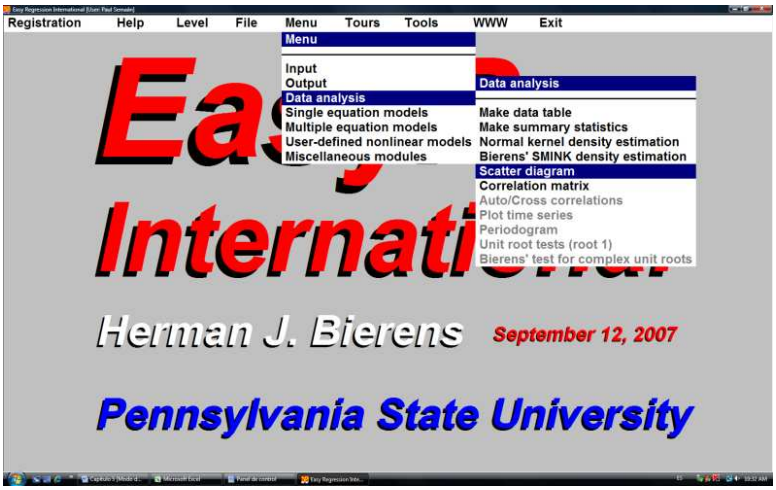
Una práctica común para detectar intuitivamente problemas de heteroscedasticidad en el término de error es emplear gráficas de dispersión de los errores estimados ($\hat{\varepsilon}$) y de las variables independientes, al igual que los errores estimados y el valor estimado de la variable dependiente. Algunos autores también sugieren emplear gráficos de dispersión del cuadrado de los residuos ($\hat{\varepsilon}^2$) y las variables explicativas. Estos gráficos son empleados para determinar la existencia de algún patrón en la variabilidad del error. En este caso solo graficaremos los errores contra las variables explicatorias $yedu_i$, exp_i, exp_i^2 y D_i

El primer paso será generar la serie de los residuos, para esto cargue los datos y corra la regresión correspondiente al modelo 5.6 empleando el método de mínimos cuadrados ordinarios.[10] Haga clic en el menú "Options" y escoja la opción "Write residuals to the input file" (primera opción), esto creará una nueva variable con el nombre "OLS Residuals of Lnih". Ahora regrese a la ventana principal de EasyReg (Haga clic la opción "Done" de la parte superior izquierda)

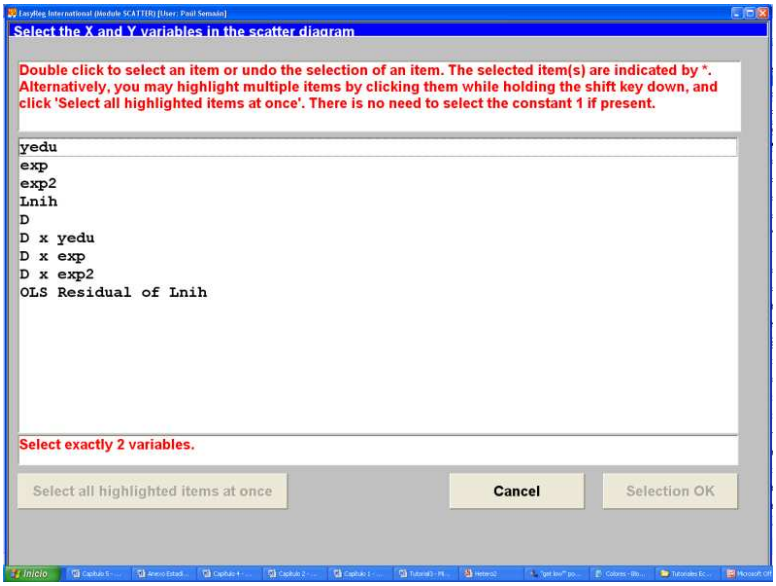


Ahora grafiquemos los residuos versus los años de experiencia. Para esto haga clic en "Menu /Data analisys / Scatter diagram"

[10] Para ver detalles de cómo hacer esto revise el capítulo 1.

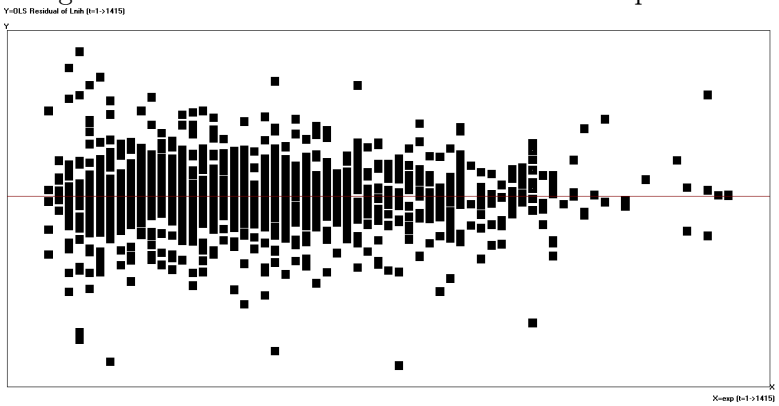


Verá la siguiente ventana:



Seleccione la variable "exp" (haciendo doble clic sobre ella) y posteriormente seleccione la variable "OLS Residuals of Lnih". Note que la primera variable seleccionada será ubicada en el eje de las $X's$ y la segunda se ubicará en el eje de las y´s. La siguiente ventana le preguntará si desea graficar una sub-muestra de los datos, haga clic en "No" y después en "Continue". Observará el siguiente gráfico.

Figura 5.2: Errores estimados versus años de experiencia

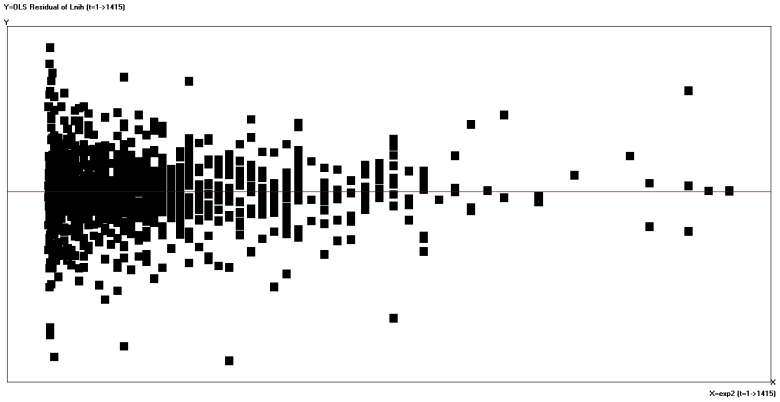


El gráfico muestra una relación no muy clara entre la variabilidad de los errores y los años de experiencia.

En todos los gráficos siempre hay una opción de grabar el gráfico dibujado por EasyReg si se hace clic en el botón "Save Picture". El gráfico queda grabado como un archivo .bmp en el fólder C:\Archivos de Programas\EasyReg International\TEMP\EASYREG.DAT o en su defecto en el fólder donde usted halla inicializado EasyReg. Los gráficos se pueden consultar rápidamente por el menú "Menu/Output/View bitmap files". Si usted está escribiendo un informe y desea pegar este gráfico en un archivo de Word, lo puede hacer fácilmente empleando el menú "Insertar / Imagen de un archivo" o simplemente seleccione el gráfico desde su carpeta de origen y arrástrelo al lugar del documento donde lo desea ubicar.

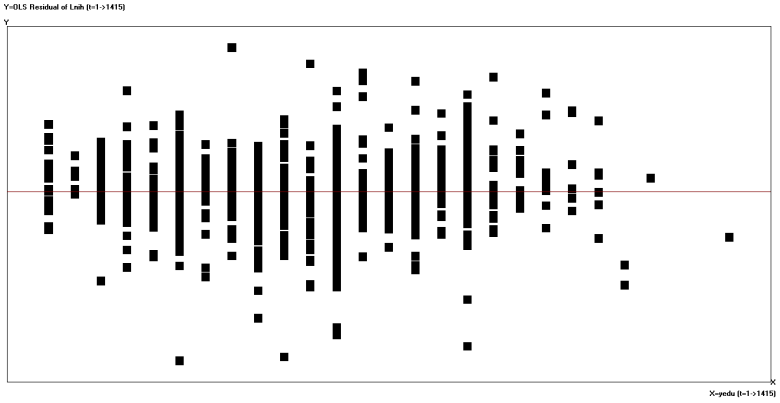
Para regresar al menú anterior haga clic en el botón "Done". Ahora, repita el mismo proceso para graficar los residuos versus la experiencia al cuadrado y las demás variables explicatorias. Obtendrá los siguientes gráficos:

Figura 5.3: Errores estimados versus años de experiencia al cuadrado



En el anterior gráfico vemos que la dispersión de los residuos disminuye a medida que se incrementa el cuadrado de la experiencia. Este es un síntoma de heteroscedasticidad

Figura 5.4: Errores estimados versus años de educación



En la figura 5.2 no podemos apreciar una relación clara entre la variabilidad de los errores y los años de educación.

5.4.2. Pruebas de heteroscedasticidad

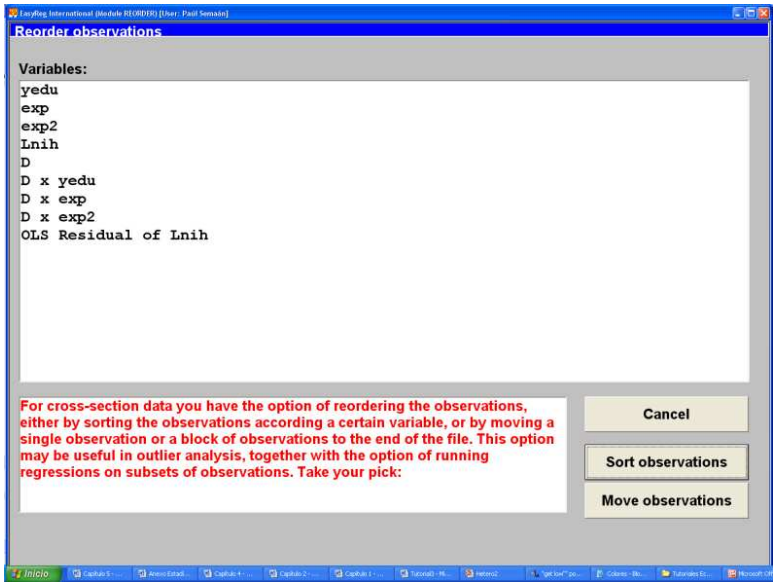
Prueba de Goldfeld y Quandt

Como ya lo estudiamos en este capítulo, esta prueba posee como hipótesis nula la presencia de homoscedasticidad versus la hipótesis alterna que $\sigma_i^2 = \sigma^2 X_{p,i}^2$, donde X_p corresponde a una variable que se sospecha está creando el problema de heteroscedasticidad.

Así, el primer paso es ordenar los datos de menor a mayor, de acuerdo a la variable que creamos esté causando el problema de heteroscedasticidad, en este caso los años de experiencia al cuadrado (exp^2). Esto se puede realizar fácilmente en EasyReg con la opción "Menu / Input / Reorder cross-section

observations". Es importante anotar que cuando se trabaja con series de tiempo, esta opción no estará disponible. En este caso usted necesitará organizar los datos con Excel.

Una vez usted escoja esta opción verá la siguiente ventana:

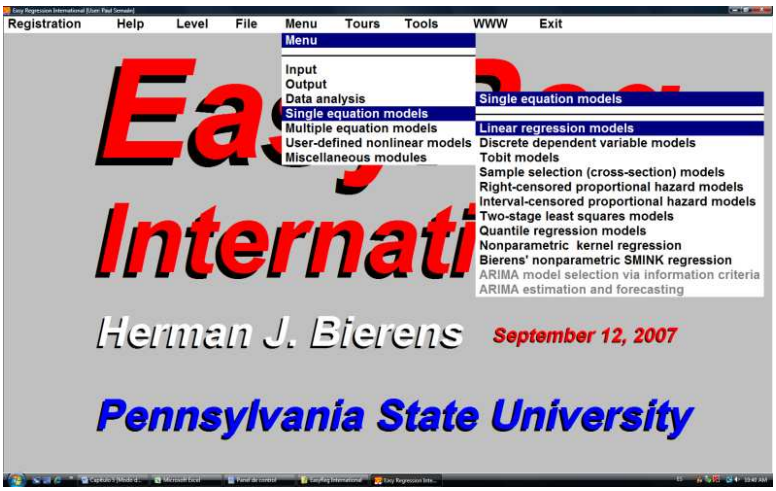


Haga clic en el botón "Sort observations" y escoja la variable que va a servir como criterio para ordenar los datos, es decir, haga clic en "exp2" y posteriormente en el botón "Sort observations". En este momento los datos ya se encuentran organizados de mayor a menor de acuerdo a la variable exp y listos para la estimación. Ahora haga clic en el botón "Overwrite input file" y después en el botón "Done". Los datos ya han sido organizados de menor a mayor

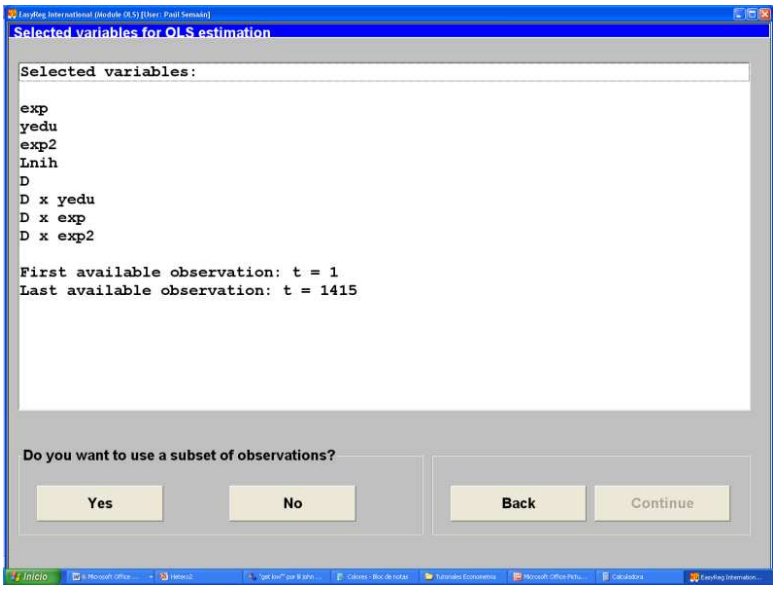
Ahora necesitamos correr dos regresiones, la primera con los datos asociados a los valores bajos de exp_i^2 y la segunda con los valores altos de exp_i^2, excluyendo d observaciones de la mitad. Esto se puede hacer fácilmente en EasyReg. Primero determine cuáles van a ser las observaciones que corresponderán a la primera y segunda regresión, es decir, emplearemos las observaciones desde $i = 1, 2, ..n_1$ para la primera regresión y las observaciones $i = n_1 + d, n_1 + d + 1, n_1 + d + 2, ...n$ en la segunda regresión. En este caso eliminaremos 281 datos de la mitad.

Así, nuestra primera regresión incluirá de la observación (ordenada) número 1 a la 567 y la segunda regresión incluirá de la observación 848 a la 1415. Ahora corramos cada una de estas regresiones y extraigamos el SSR reportado por EasyReg de cada una de las regresiones.

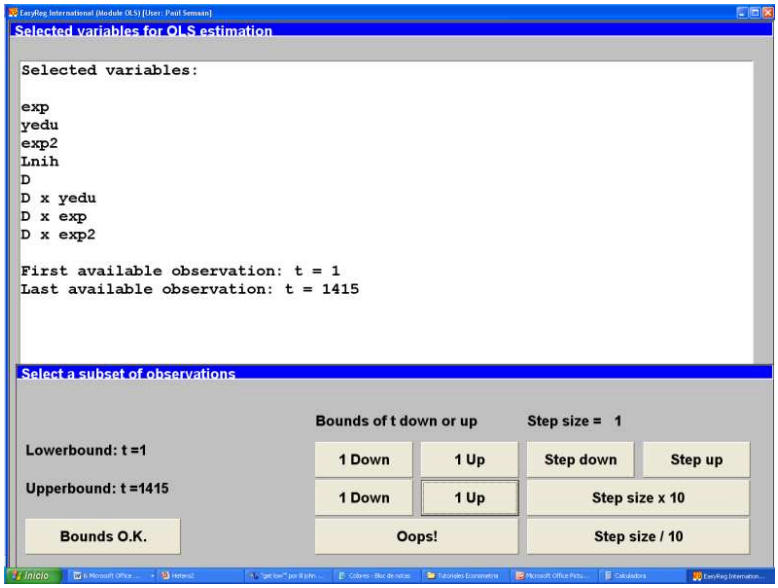
Para correr la primera regresión, haga clic en "Menu / Single equation models / linear regression models".



Ahora, escoja las variables involucradas en la regresión y haga clic en el botón "Selection OK". En la siguiente ventana haga clic en el botón "Yes" (Esto le permitirá escoger la sub-muestra que quiere emplear para la estimación del modelo).



Observará la siguiente ventana:



Con los botones "1 Down" y "1 Up" escoja la muestra, de tal forma que el "Lowerbound: t=1" y el "Upperbound: t=567". Posteriormente, haga clic en "Bounds OK". Ahora haga clic en "Continue", después en "Confirm" y nuevamente en "Continue". Ahora prosiga con los pasos que usted realiza normalmente para la estimación de un modelo. Encontrará que el SSE_2=194.35.

Repita los pasos anteriores y corra la regresión con las observaciones asociadas con los valores altos de x_i, es decir, de la observación 848 a la 1415. Encontrará que SSE_1=186.77. Ahora podemos calcular el estadístico:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{194{,}35}{186{,}77} = 1{,}04$$

Este F_{GQ} lo comparamos con:

$$F_{(1118,1118),\alpha=0,1} = 1{,}079$$

Dado que $F_{GQ}<F_{(1118,1118),\alpha=0,1}$, no podemos rechazar la hipótesis nula de varianza homoscedastica. Al realizar este mismo procedimiento ordenando los datos según los años de educación obtenemos que:

$$F_{GQ} = \frac{SSE_2}{SSE_1} = \frac{194{,}87}{151{,}36} = 1{,}287$$

Al compararlo con $F_{(1118,1118),\alpha=0,01} = 1{,}149$, tenemos que $F_{GQ}>F_{(1118,1118),\alpha=0,1}$ y por lo tanto podemos rechazar la hipótesis nula a un nivel de significancia del 1 %. Estos resultados se resumen en el cuadro 5.1. Este resultado concuerda con lo observado en la figura 5.3.

Cuadro 5.1: Resultados de la Prueba de Goldfeld y Quandt

	Nombre de la variable			
	exp		*yedu*	
SSE_1	186.770		151.360	
SSE_2	194.350		194.870	
F_{GQ}	1.041	*	1.287	***

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

Prueba de Breusch-Pagan

Es importante resaltar que EasyReg calcula el estadístico de Breush Pagan que corresponde a la hipótesis alterna que todas las variables incluidas en la regresión causan el problema de Heteroscedasticidad. Por ejemplo, al estimar el modelo (5.6) se obtiene el estadístico BP de 53.05 que corresponde a la hipótesis alterna:

$$\sigma_i^2 = f(yedu, exp, exp^2, D, D(yedu), D(exp), D(exp^2))$$

Ese estadístico sigue una distribución ^{2} con siete grados de libertad. Así, en este caso se puede rechazar la H_o de homoscedasticidad a un nivel de significancia del 5 %. De ser necesario comprobar otra hipótesis alterna, entonces se puede realizar los siguientes pasos.

Note que previamente hemos creado la serie de los residuos en EasyReg, ahora necesitamos encontrar:

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n}$$

Para efectuar este cálculo, podemos emplear los resultados arrojados por la regresión inicial. Recuerde que $\hat{\varepsilon}^T\hat{\varepsilon}$=Suma de los Errores de la Regresoón $(SST - SSRN)$ de esta forma tenemos que:

$$\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{n} = \frac{463{,}87}{1415} = 0{,}328$$

Anteriormente habíamos creado la serie de los residuos, ahora sólo necesitamos elevarla al cuadrado y posteriormente dividir esa serie por 0.328 (esto se puede realizar al emplear la opción de "Multiplicative combination of variables" para elevar la serie al cuadrado y luego mediante "linear combination of variables" (en el menú de "Transform variables") y multiplicar por 1/0.2731=3.653 para obtener la serie que será empleada como variable dependiente en el siguiente paso. Cree esta nueva variable y llámela $ERBP$. Ahora corra la regresión auxiliar:

$$\frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta Z_i + \mu_i$$

Para continuar con nuestro ejemplo, efectuaremos la prueba de Goldfeld y Quandt para las H_A presentadas en el cuadro 5.2

Cuadro 5.2: Hipótesis alternas de la prueba de Breush Pagan

H_A	Regresión auxiliar
σ^2=$f(exp^2)$	$\text{ERPB}_i = \beta_0 + \beta_1 \text{exp}_i^2 + \varepsilon$
σ^2=$f(exp)$	$\text{ERPB}_i = \beta_0 + \beta_1 \text{exp}_i + \varepsilon$
σ^2=$f(yedu)$	$\text{ERPB}_i = \beta_0 + \beta_1 yedu_i + \varepsilon$
σ^2=$f(D)$	$\text{ERPB}_i = \beta_0 + \beta_1 D_i + \varepsilon$
σ^2=$f(D, yedu, exp, exp^2)$	$\text{ERPB}_i = \beta_0 + \beta_1 \text{exp}_i^2 + \beta_2 \text{exp}_i + \beta_3 yedu_i + \beta_4 D_i + \varepsilon$

En cada estimación utilizamos el SSE y el SST para calcular el SSR ($SSR = SST - SSE$). Posteriormente calculamos el estadístico de Breush Pagan (BP) mediante la división $SSR/2$ y este valor lo contrastamos con el estadístico chi-cuadrado con g grados de libertad (donde g corresponde al número de variables explicativas incluidas en cada uno de los modelos estimados). Los resultados se resumen en el cuadro 5.3.

Cuadro 5.3: Resultados de la prueba de Breush Pagan

	Hipótesis alterna				
	σ^2=$f(exp^2)$	σ^2=$f(exp)$	σ^2=$f(yedu)$	σ^2=$f(D)$	σ^2=$f(D, yedu, exp, exp^2)$
SST	7616.37	7616.37	7616.37	7616.37	7616.37
SSE	7616.27	7611.79	7572.54	7609.42	7528.08
BP	0.046	2.288	21.92***	3.47*	44.15***

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

Por lo tanto los resultados de esta prueba indican que dos de las variables explicativas están causando problemas de heteroscedasticidad.

Prueba de White

Por último se realiza la prueba de White, para contrastar la hipótesis nula de no heteroscedasticidad versus la alterna de heteroscedasticidad. Para llevar a cabo esta prueba hay que estimar una regresión auxiliar donde la variable dependiente sea igual a los residuos al cuadrado de la formulación inicial y las variables independientes sean conformadas por regresores del modelo original, por sus cuadrados y por los productos cruzados, después de haber eliminado las posibles redundancias se obtiene el siguiente modelo auxiliar a estimar:

$$\begin{aligned}\hat{\varepsilon}_i^2 =& \beta_0 + \beta_1 edu_i + \beta_2 ex_i + \beta_3 ex_i^2 + \beta_4 D_i + \beta_5 D_i edu_i + \beta_6 D_i ex_i + \beta_7 D_i ex_i^2 \\ &+ \beta_8 edu_i^2 + \beta_9 ex_i^4 + \beta_{10} D_i edu_i^2 + \beta_{11} D_i ex_i^4 + \beta_{12} edu_i ex_i + \beta_{13} edu_i ex_i^2 \\ &+ \beta_{14} D_i edu_i ex_i + \beta_{15} ex_i^3 + \beta_{16} D_i edu_i ex_i + \beta_{17} ex_i^3 + \varepsilon_i\end{aligned} \tag{5.7}$$

El R^2 obtenido a partir de la regresión auxiliar es igual a 0.0274. Ahora se puede construir el

estadístico de White: $W_a = nR^2 = 1415 \times 0{,}0274 = 38{,}77$ el cual se debe comparar con $\chi^2_{g(\alpha)} = \chi^2_{17(0{,}01)} = 33{,}41$. De acuerdo a los valores críticos obtenidos, la hipótesis nula se puede rechazar a este nivel de significancia.

Por lo tanto esta prueba, al igual que las dos anteriormente efectuadas, indica la presencia de heteroscedasticidad en el modelo. A continuación estudiaremos la forma más adecuada de resolver este problema.

5.4.3. Solución al problema de heteroscedasticidad

Las pruebas anteriormente realizadas muestran claramente que el modelo presenta problemas de heteroscedasticidad, sin embargo estas mismas pruebas indican que el problema está siendo causado por más de una variable y no sabemos la forma funcional. Esto imposibilita que podamos determinar la forma exacta de la varianza de los errores y por lo tanto el método de Mínimos Cuadrados Ponderados (MCP) no es viable en este caso, es por esto que debemos solucionar el problema empleando la solución propuesta por White.

Recuerden que cuando estimamos un modelo por el método de MCO en presencia de heteroscedasticidad, entonces los estimadores de los $\beta's$ son insesgados, pero el estimador de la matriz de varianzas y covarianzas de los $\beta's$ estimados es sesgado. White (1980) sugiere el siguiente estimador consistente para la matriz de varianzas y covarianzas de los $\beta's$ estimados en presencia de heteroscedasticidad:

$$Est.Var\left[\hat{\beta}\right] = n\big(X^TX\big)^{-1}S_0\big(X^TX\big)^{-1} \tag{5.8}$$

donde

$$S_0 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2 x_i x_i^T \quad \text{y} \quad x_i^T = \begin{pmatrix} 1 & x_{1i} & x_{2i} & \dots & x_{ki} \end{pmatrix} \tag{5.9}$$

Cuando estimamos una regresión lineal, EasyReg calcula automáticamente esta matriz, exista o no problema de heteroscedasticidad. Así que ustedes deben tener cuidado para saber cuándo se debe emplearla o no.

Estime el modelo original:

$$\begin{aligned}\ln(ih_i) =& \beta_0 + \beta_1 yedu_i + \beta_2 exp_i + \beta_3 exp_i^2 + \beta_4 D_i \\ &+\beta_5 Dyedu_i + \beta_6 Dexp_i + \beta_7 Dexp_i^2 + \varepsilon_i\end{aligned} \tag{5.10}$$

donde

$$D_t = \begin{cases} 1 & \text{si el individuo } i \text{ es hombre} \\ 0 & o.w. \end{cases}$$

Obtendrá los resultados del siguiente recuadro:

Recuadro 5.2 Output de EasyReg modelo 5.10

Dependent variable:
Y = Lnih
Characteristics:
Lnih
First observation = 1
Last observation = 1415
Number of usable observations: 1415
Minimum value: 4.0173836E+000
Maximum value: 1.0750786E+001
Sample mean: 7.5250474E+000

X variables:
X(1) = yedu
X(2) = exp
X(3) = exp2
X(4) = D
X(5) = D x yedu
X(6) = D x exp
X(7) = D x exp2
X(8) = 1

Model:
Y = b(1)X(1) +.....+ b(8)X(8) + U,
where U is the error term, satisfying
E[U|X(1),...,X(8)] = 0.
OLS estimation results

Parameters	Estimate	t-value (S.E.) [p-value]	H.C. t-value (H.C. S.E.) [H.C. p-value]
b(1)	0.13653	23.743 (0.00575) [0.00000]	21.688 (0.00629) [0.00000]
b(2)	0.03048	5.407 (0.00564) [0.00000]	5.210 (0.00585) [0.00000]
b(3)	-0.00023	-1.990 (0.00012) [0.04660]	-2.041 (0.00011) [0.04123]
b(4)	0.24800	1.983 (0.12504) [0.04733]	1.709 (0.14508) [0.08737]
b(5)	-0.01131	-1.420 (0.00797) [0.15572]	-1.257 (0.00900) [0.20867]
b(6)	0.00339	0.433 (0.00784) [0.66519]	0.383 (0.00886) [0.70152]

b(7)	-0.00008	-0.495	-0.452
		(0.00015)	(0.00017)
		[0.62086]	[0.65136]
b(8)	5.66732	62.633	55.859
		(0.09048)	(0.10146)
		[0.00000]	[0.00000]

Notes:
1: S.E. = Standar* error
2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
3: The two-sided p-values are based on the normal approximation.

Effective sample size (n): 1415
Variance of the residuals: 0.329689
Standard error of the residuals (SER): 0.574185
Residual sum of squares (RSS): 463.872119
(Also called SSR = Sum of Squared Residuals)
Total sum of squares (TSS): 851.238679
R-square: 0.4551
Adjusted R-square: 0.4524

Overall F test: F(7,1407) = 167.85
p-value = 0.00000

Significance levels:	10 %	5 %
Critical values:	1.72	2.01
Conclusions:	reject	reject

If the model is correctly specified, in the sense that the conditional expectation of the model error U relative to the X variables equals zero, then the OLS parameter estimators b(1),..,b(8), minus their true values, times the square root of the sample size n, are (asymptotically) jointly normally distributed with zero mean vector and variance matrix:

4.68E-02	3.76E-03	9.71E-05	5.70E-01	-4.68E-02	-3.76E-03	-9.71E-05	-5.70E-01
3.76E-03	4.50E-02	-8.63E-04	4.44E-01	-3.76E-03	-4.50E-02	8.63E-04	-4.44E-01
9.71E-05	-8.63E-04	1.90E-05	-5.68E-03	-9.71E-05	8.63E-04	-1.90E-05	5.68E-03
5.70E-01	4.44E-01	-5.68E-03	2.21E+01	-1.04E+00	-8.72E-01	1.14E-02	-1.16E+01
-4.68E-02	-3.76E-03	-9.71E-05	-1.04E+00	8.99E-02	4.32E-03	1.87E-04	5.70E-01
-3.76E-03	-4.50E-02	8.63E-04	-8.72E-01	4.32E-03	8.71E-02	-1.61E-03	4.44E-01
-9.71E-05	8.63E-04	-1.90E-05	1.14E-02	1.87E-04	-1.61E-03	3.38E-05	-5.68E-03
-5.70E-01	-4.44E-01	5.68E-03	-1.16E+01	5.70E-01	4.44E-01	-5.68E-03	1.16E+01

provided that the conditional variance of the model error U is constant (U is homoskedastic), or

5.61E-02	9.93E-03	2.27E-05	7.19E-01	-5.61E-02	-9.93E-03	-2.27E-05	-7.19E-01
9.93E-03	4.84E-02	-8.72E-04	5.77E-01	-9.93E-03	-4.84E-02	8.72E-04	-5.77E-01
2.27E-05	-8.72E-04	1.80E-05	-7.34E-03	-2.27E-05	8.72E-04	-1.80E-05	7.34E-03
7.19E-01	5.77E-01	-7.34E-03	2.98E+01	-1.37E+00	-1.29E+00	1.75E-02	-1.46E+01
-5.61E-02	-9.93E-03	-2.27E-05	-1.37E+00	1.15E-01	1.54E-02	6.48E-05	7.19E-01
-9.93E-03	-4.84E-02	8.72E-04	-1.29E+00	1.54E-02	1.11E-01	-1.99E-03	5.77E-01
-2.27E-05	8.72E-04	-1.80E-05	1.75E-02	6.48E-05	-1.99E-03	4.06E-05	-7.34E-03
-7.19E-01	-5.77E-01	7.34E-03	-1.46E+01	7.19E-01	5.77E-01	-7.34E-03	1.46E+01

if the conditional variance of the model error U is not constant (U is heteroskedastic).

La última matriz que ustedes observan en los resultados de EasyReg corresponde al estimador

consistente de White para la matriz de varianzas y covarianzas de los $\beta's$ estimados multiplicada por el número de observaciones.[11] Basados en esta matriz, podemos re-calcular los t-calculados, de tal forma que tengan en cuenta el estimador consistente de las varianzas de los $\beta's$ estimados. Esto también lo hace automáticamente EasyReg. Estos t-calculados corresponden a la columna con encabezado "H.C. t-value(*)" y su correspondiente "p-value" se encuentra debajo entre corchetes. En el cuadro 5.4 se presentan los resultados de la estimación, la cual muestra que existe un diferencial positivo y significativo de 0.248 en el término que recoge el valor medio del logaritmo natural del ingreso por hora cuando el género del individuo corresponde a "hombre". Por lo tanto, podemos afirmar que el género es un causante en las diferencias salariales en Cali a pesar de que no existe diferencia entre hombres y mujeres en el impacto que generan los años de educación o de experiencia sobre el logaritmo natural del ingreso.[12]

[11] Para ser más exactos, esa matriz se debe multiplicar por $\frac{1}{n}$ para obtener el valor estimado para 5.8.

[12] Para poder afirmar esto último se realizó una última prueba de Wald de significancia conjunta de los coeficientes que acompañan a las variables $D_i yedu_i$, $D_i exp_i$ y $D_i exp_i^2$. El estadístico Wald para esta prueba es de 2.05, este estadístico sigue una distribución chi-cuadrado con 3 gardos de libertad. Dado que los valores críticos con niveles de significancia del 10 % y 5 % son respectivamente 6.25 y 7.81 respectivamente entonces no es posible rechazar la hipótesis nula de que estos coeficientes son conjuntamente iguales a cero.

Cuadro 5.4: Estimación del modelo 5.10

	Variable dependiente $lnih_i$ Estadísticos t entre paréntesis Basado en la matriz de varianzas y covarianzas consistente de White
	Ecuación 5.10 MCO
Constante	5.66732 (55.859)***
$yedu_i$	0.13653 (21.688)***
$expi$	0.03048 (5.21)***
exp_i^2	-0.00023 (-2.041)**
D_i	0.248 (1.709)*
$D_i yedu_i$	-0.01131 (-1.257)
$D_i exp_i$	0.00339 (0.383)
$D_i expi^2$	-0.0000765 (-0.452)
R^2	0.45510
R2 Ajustado	0.45240
F-Global	167.85***
No. de obs.	1415

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos Cuadrados Ordinarios

5.5. Ejercicios

Una cadena de Supermercados quiere planear el número de almacenes a inaugurar en España en los próximos años. Para ello desea conocer la relación que existe entre el consumo final de las familias residentes y la renta bruta de las mismas. Usted ha sido contratado como consultor y cuenta con una base de datos para 18 regiones (los datos se encuentran en el archivo het.xls). Ambas variables están medidas en millones de pesetas. El modelo es el siguiente:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i \tag{5.11}$$

Donde Y_i representa el consumo final de las familias para la región i, y X_i representa la renta bruta de las familias para la región (i).

1. Estime el modelo 5.11

2. Realice las pruebas estudiadas en este capítulo para determinar si el problema de heteroscedasticidad se encuentra presente en el modelo, plantee las hipótesis nula y alterna de cada prueba.

3. Si existe heteroscedasticidad, resuelva el problema y estime un modelo que no presente este problema. Demuestre teóricamente que el problema ha desaparecido.

4. Interprete los coeficientes estimados.

5.6. Apéndice

Apéndice 5.1 Demostración de la insesgadez de los estimadores en presencia de heteroscedasticidad

En presencia de heteroscedasticidad los estimadores MCO siguen siendo insesgados. Esta afirmación se puede demostrar fácilmente. Consideremos un modelo lineal con un término de error heteroscedástico y no autocorrelación. Es decir,

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

Donde $E[\varepsilon_t] = 0$, $Var[\varepsilon_t] = \sigma_\varepsilon^2$ y $E[\varepsilon_i \varepsilon_j] = 0$, $\forall i \neq j$. Ahora determinemos si $\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$ sigue siendo insesgado o no. Así,

$$E\left[\hat{\beta}\right] = E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} E[\mathbf{y}]$$

$$E\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} E[\mathbf{X}\beta + \varepsilon] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta + \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T} E[\varepsilon]$$

$$E\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta = I \bullet \beta$$

$$E\left[\hat{\beta}\right] = \beta$$

Apéndice 5.2 Sesgo de la matriz de varianzas y covarianzas en presencia de heteroscedasticidad

En presencia de heteroscedasticidad el estimador de la matriz de varianzas y covarianzas de MCO ($\widehat{Var\left[\hat{\beta}\right]} = s^2\left(\mathbf{X^T X}\right)$) es sesgado. Es más, el estimador MCO para los coeficientes ($\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$) no es eficiente; es decir no tiene la mínima varianza posible. Esta afirmación se puede demostrar fácilmente.

Continuando con el modelo considerado en el Apéndice anterior, en este caso tenemos que:

$$Var[\varepsilon] = E\left[\varepsilon^T \varepsilon\right] = \Omega = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & \sigma_n^2 \end{bmatrix} \tag{5.12}$$

Ahora podemos calcular la varianza de los estimadores MCO. Es decir,

$$Var\left[\hat{\beta}\right] = Var\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}\right] \tag{5.13}$$

.

Por tanto tendremos que

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T} Var\left[\mathbf{y}\right] \mathbf{X} \left(\mathbf{X^T X}\right)^{-1} \tag{5.14}$$

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T} Var\left[\varepsilon\right] \mathbf{X} \left(\mathbf{X^T X}\right)^{-1} \tag{5.15}$$

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T} \Omega \mathbf{X} \left(\mathbf{X^T X}\right)^{-1} \tag{5.16}$$

Por tanto la varianza no es la mínima posible. Y por otro lado, al emplear el estimador MCO para la matriz de varianzas y covarianzas de los betas ($\widehat{Var\left[\hat{\beta}\right]} = s^2 \left(\mathbf{X^T X}\right)^{-1}$) en presencia de heteroscedasticidad se obtendrá un estimador cuyo valor esperado no es igual a la varianza real; es decir, será insesgado.

CAPÍTULO 6

Autocorrelación

Objetivos del capítulo

Al final de este capítulo el lector estará en condiciones de:

- Realizar en EasyReg diferentes tipos de análisis gráficos que revelen la posibilidad de autocorrelación en los residuos.
- Efectuar mediante EasyReg las pruebas estadísticas necesarias para detectar la violación del supuesto de no autocorrelación en los residuos. En especial las pruebas de Rachas, Durbin Watson, Box-Pierce y de Ljung-Box.
- Solucionar por medio de EasyReg el problema de Autocorrelación.

6.1. Introducción

En los dos capítulos anteriores hemos analizado las consecuencias de violar algunos de los supuestos del modelo de regresión. En el Capítulo 4 analizamos las consecuencias de la violación del supuesto de independencia lineal entre variables explicativas en un modelo de regresión múltiple, posteriormente en el Capítulo 5 nos enfocamos en las consecuencias de un error homoscedástico.

Finalmente, y para concluir nuestra discusión de la violación de los supuestos que garantizan el cumplimiento del Teorema de Gauss-Markov (Ver recuadro 6.1) nos concentraremos en los efectos que tiene la violación del supuesto de no autocorrelación (no existe relación entre los diferentes errores). Este supuesto garantiza la no presencia de un patrón predecible en el comportamiento de los errores. Cuando este supuesto es violado, lo cual ocurre comúnmente cuando trabajamos con datos de series de tiempo, se dice que los errores presentan autocorrelación (o correlación serial), en otras palabras, están relacionados entre sí.

Recuadro 6.1 Teorema de Gauss-Markov

Si se considera un modelo lineal $\mathbf{y_{n\times 1}} = \mathbf{X_{n\times k}}\beta_{k\times 1} + \varepsilon_{n\times 1}$ y se supone que:

- Las $X_2, X_3, ..., X_k$ son fijas y linealmente independientes (es decir X tiene rango completo y es una matriz no estocástica).
- El vector de errores ε tiene media cero, varianza constante y no autocorrelación. Es decir: $E[\varepsilon] = 0$ y $Var[\varepsilon] = \sigma^2 I_n$

Entonces el estimador de MCO $\hat{\beta} = \left(X^T X\right)^{-1} X^T \mathbf{y}$ es el *Mejor Estimador Lineal Insesgado (MELI)*

Si existe autocorrelación entre los errores, entonces los estimadores MCO siguen siendo insesgados pero no son eficientes (ver 3 y 4 para la demostración de estos resultados).

Veamos más en detalle qué significa la autocorrelación. Y para simplificar, estudiemos inicialmente el caso más sencillo. Cuando existe una relación lineal "grande" entre las observaciones adyacentes, pero esta relación (lineal) tiende a desaparecer a medida que se consideran errores más lejanos. Formalmente tenemos:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \tag{6.1}$$

donde el término del error tiene media cero y se encuentra correlacionado con el error del periodo anterior: $E[\varepsilon_t] = 0$; $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t \quad \forall t$ con $0 \leq |\rho| < 1$ y $Var(v_t) = \sigma_v^2$.

Este tipo de autocorrelación es conocido como un proceso auto-regresivo de orden uno o AR(1) para abreviar. Si seguimos asumiendo que la varianza de los errores es constante, entonces se puede probar fácilmente (hágalo) que:

$$\sigma_\varepsilon^2 = \frac{\sigma_\nu^2}{(1-\rho^2)}$$

Por otro lado, dado que el valor esperado del error es cero se puede mostrar fácilmente (hágalo):

$$Cov(\varepsilon_t, \varepsilon_{t-1}) = \rho\sigma_\varepsilon^2$$

Entonces, la correlación entre los errores adyacentes[1] será:

$$\frac{Cov\left(\varepsilon_t, \varepsilon_{t-1}\right)}{\sqrt{Var\left(\varepsilon_t\right) Var\left(\varepsilon_{t-1}\right)}} = \frac{Cov\left(\varepsilon_t, \varepsilon_{t-1}\right)}{\sigma_\varepsilon^2} = \rho$$

Similarmente es relativamente sencillo demostrar que:

$$Cov\left(\varepsilon_t, \varepsilon_{t-2}\right) = \rho^2 \sigma_\varepsilon^2$$

$$Cov\left(\varepsilon_t, \varepsilon_{t-3}\right) = \rho^3 \sigma_\varepsilon^2$$

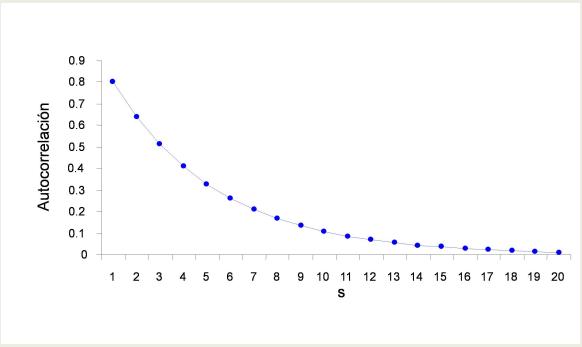
En términos generales teneos que en este caso de errores con un proceso AR(1) la autocorrelación para diferentes rezagos está dada por:[2]

$$\rho\left(s\right) = \rho^s \sigma_\varepsilon^2$$

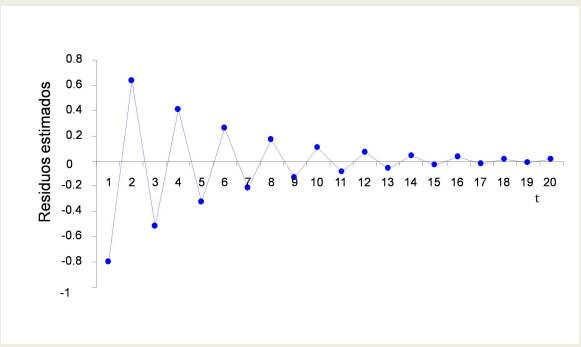
Es decir, en el caso de un proceso AR(1) a medida que se alejan en el tiempo las observaciones (se consideran más rezagos) la correlación entre los errores es menor. Esto lo podemos observar en el ejemplo 6.1. Si no existiese un problema de autocorrelación, la autocorrelación para los diferentes rezagos será cero.

Ejemplo 6.1 Autocorrelación de los errores con un proceso AR(1)

ρ=0.8



ρ=-0.8



Regresando al problema de autocorrelación, su origen puede ser causado porque las relaciones entre las variables pueden ser dinámicas. Es decir, todo el ajuste entre las variables no se hace en un mismo período; un ejemplo de este comportamiento son las expectativas adaptativas. La autocorrelación también se puede deber a que la información no está disponible instantáneamente y por tanto, la variable dependiente puede depender de errores previos. Por ejemplo, la información del PIB no se encuentra disponible sino después de varios períodos, y los agentes tendrán que esperar unos periodos para ajustar sus decisiones. En general, la autocorrelación es un problema muy común en modelos que emplean series de tiempo.

[1] La correlación entre errores inmediatamente adyacentes, es decir separados por únicamente un periodo también se domina la autocorrelación a un rezago. Si se considera la relación entre errores separados por dos periodos se denomina autocorrelación para a dos rezagos, etc.

[2] A la función que muestra correlación para diferentes rezagos de un proceso se le denomina función de autocorrelación.

La autocorrelación entre los errores puede tomar muchas formas. En general, se dirá que los errores siguen un proceso autorregresivo de orden p ($AR(p)$) cuando el error depende de los p períodos anteriores. Por ejemplo, si el término de error tiene el siguiente comportamiento $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \nu_t$ ($\forall t$), entonces se dirá que el error es auto-regresivo de orden 2 ($AR(2)$). Si el comportamiento es $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \rho_3\varepsilon_{t-3} + \nu_t$, entonces los errores seguirán un proceso $AR(3)$. En general, si el comportamiento del error es $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \rho_3\varepsilon_{t-3} + \cdots \rho_p\varepsilon_{t-p} + \nu_t$ entonces el error sigue un proceso AR(p).

Matricialmente, la presencia de autocorrelación implica que la matriz de varianzas y covarianzas del término de error no es $\sigma^2 I_n$, sino una matriz cuadrada con la misma constante sobre la diagonal pero por fuera de la diagonal ya no se tienen ceros. Por ejemplo, en el caso de un error que sigue un proceso $AR(1)$ la siguiente matriz de varianzas y covarianzas de los errores será:

$$E\left[\varepsilon^T\varepsilon\right] = \Omega = \sigma_\varepsilon^2 \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho & 1 \end{bmatrix} \neq \sigma^2 I_n$$

Como se mencionó anteriormente, en presencia de autocorrelación los estimadores MCO continúan siendo insesgados pero no tienen la mínima varianza posible.[3] Es más, en presencia de autocorrelación, el estimador MCO de la matriz de varianzas y covarianzas del vector β será sesgado.[4] Por lo tanto, si usamos este último estimador en pruebas de hipótesis (individuales o conjuntas) o intervalos de confianza para los coeficientes estimados, entonces obtendremos conclusiones erróneas en torno a los verdaderos $\beta's$.

6.2. Pruebas para la detección de autocorrelación

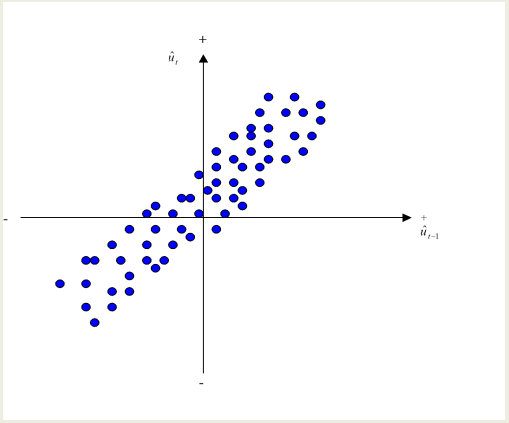
Ejemplo 6.2 Tipos de autocorrelación de primer orden

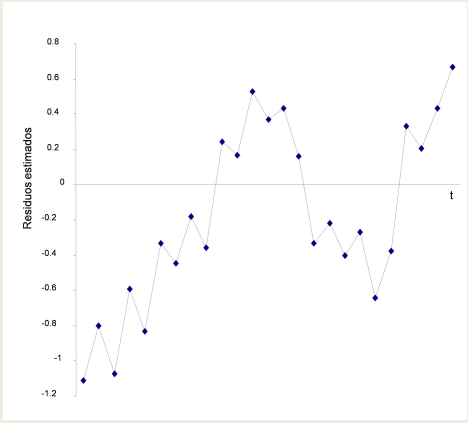
A continuación nos concentraremos en procesos $AR(1)$ por simplicidad. Cuando los errores siguen un proceso $AR(1)$, se pueden distinguir dos tipos de autocorrelación:

1. Autocorrelación positiva ($0 < \rho < 1$). En este caso, los errores de períodos adyacentes tienden a tener el mismo signo.
2. Autocorrelación negativa ($-1 < \rho < 0$). En este caso, los errores de períodos adyacentes tienden a tener diferente signo.

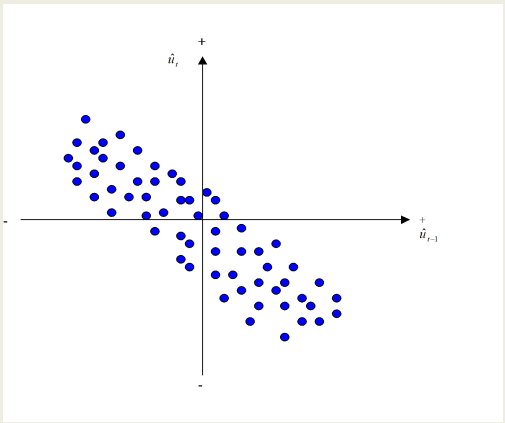
[3] En el apéndice 6.3 se presenta una demostración de esta afirmación.
[4] En el apéndice 6.4 se presenta una demostración de esta afirmación.

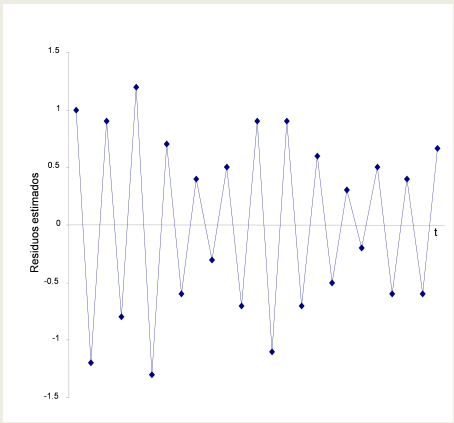
Autocorrelación positiva



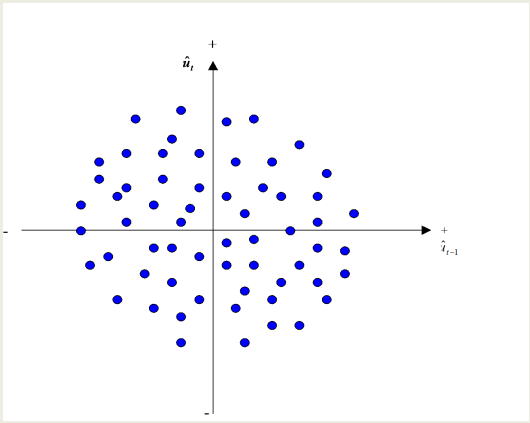


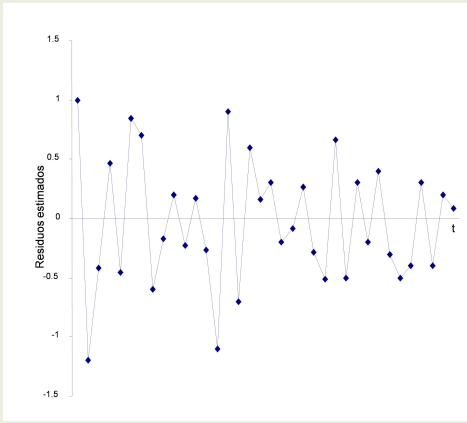
Autocorrelación negativa





No autocorrelación





El primer paso, antes de realizar pruebas estadísticas, es verificar si los síntomas de este problema están presentes. Una vez más, estos síntomas pueden ser vistos en el vector de errores. Pero dicho vector no es observable, por tanto la mejor aproximación es examinar los errores estimados. Las gráficas más comunes son:

1. Los errores contra el tiempo $\hat{\varepsilon}$ vs $t = 1, 2, ...n$

2. Los errores contra los errores del periodo anterior $\hat{\varepsilon}_t$ vs $\hat{\varepsilon}_{t-1}$

Estos gráficos permiten identificar algún tipo de regularidad como los que se presentan en el ejemplo 6.1. En los dos primeros gráficos se observa el comportamiento típico de errores con un proceso $AR(1)$ y con autocorrelación positiva, observamos que el signo de los residuos persiste en periodos prolongados de tiempo. Por otro lado en el caso de presencia de autocorrelación negativa sucede lo contrario, ya que el signo de los residuos cambia de un periodo a otro, mientras que en los gráficos de no autocorrelación, no podemos observar ningún patrón claro de comportamiento en el signo de los residuales.

Así como en el caso de la heteroscedasticidad, el análisis gráfico es un análisis informal que nos permite intuir la presencia del problema, sin embargo para un análisis más formal existen diferentes pruebas para identificar la autocorrelación.

6.2.1. Prueba de Rachas (Runs test)

La prueba de rachas, propuesta por Wald y Wolfowitz (1940), es una prueba de independencia lineal no paramétrica cuya idea es relativamente sencilla. Si no hay autocorrelación, entonces no deberían haber muchos errores seguidos con el mismo signo (autocorrelación positiva), ni tampoco muchos cambios de signo seguido (autocorrelación negativa). En otras palabras, debe existir la cantidad adecuada de cambios de signo en una serie de datos: ni muchos, ni pocos.

Esta prueba tiene además la ventaja de no necesitar suponer una distribución de los errores. Para probar la hipótesis nula de que los errores son totalmente aleatorios ($H_0 : \rho = 0$) versus la alterna de que existe algún tipo de autocorrelación en los errores, se requiere seguir los siguientes pasos a partir de los errores estimados:

1. Cuente el número de errores con signo positivo (N_+) y con signo (N_-)

2. Cuente el número rachas (k), es decir de "seguidillas"de signo. Por ejemplo, si tenemos que los signos de los errores son: - - - - - + + + - +++ - ++. Entonces se tendrán seis rachas (k =6). (- - - - -)(+ + +)(-)(+++)(-)(++). Note que el número de rachas es igual al número de cambios de signo

3. Si N_+ y/o N_- son menores que 20, entonces se puede construir un intervalo de confianza del 95 % para el número de rachas "razonable" bajo la hipótesis nula a partir de los valores críticos que se presentan en el apéndice 6.1 que se presentan al final de este capítulo

4. Si N_+ y/o N_- son mayores que 20, entonces se puede construir un intervalo de confianza del $(1-\alpha)100$ % para el número de rachas "razonable" bajo la hipótesis de la siguiente manera:

$$\left[E\left[k\right] \pm z_{\frac{\alpha}{2}} \sqrt{Var\left[k\right]}\right] \tag{6.2}$$

donde, el valor esperado y la varianza de k (las rachas) son:

$$E\left(k\right) = \frac{2N_+N_-}{N_+ + N_-} + 1 \tag{6.3}$$

$$Var\left(k\right) = \frac{2N_+N_-\left(2N_+N_- - N_+ - N_-\right)}{\left(N_+ + N_-\right)^2\left(N_+ + N_- - 1\right)} \tag{6.4}$$

La hipótesis nula se puede rechazar si el número de rachas observadas no están contenidas en el intervalo de confianza.[5]

En la página web del libro se encuentra disponible un archivo de Excel que permite realizar esta prueba de manera automática, siempre y cuando N_+ y N_- sean mayores a 20.

6.2.2. Prueba de Durbin-Watson

Durbin y Watson (1951) diseñaron una prueba de autocorrelación con gran poder para detectar errores con autocorrelación de primer orden. Esta prueba se ha convertido en la más común para detectar este problema por ser relativamente intuitiva. Los autores definen el siguiente estadístico de prueba a partir de los errores estimados:

$$DW = \frac{\sum_{t=2}^{n} \left(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1}\right)^2}{\hat{\varepsilon}^T \hat{\varepsilon}}$$

Si la muestra es lo suficientemente grande es posible demostrar que $DW \approx 2\,(1-\hat{\rho})$. De esta expresión se puede deducir que este estadístico estará acotado entre cero y 4 ($0 \leq DW \leq 4$). De hecho, como se muestra en el recuadro 6.2, intuitivamente se puede conocer el tipo de problema presente en la regresión a partir del valor del estadístico DW. Naturalmente, será necesario efectuar una prueba formal para determinar con mayor certeza si existe o no autocorrelación.

Recuadro 6.2 Estadístico DW en casos de Autocorrelación

No correlación	$\rho = 0$	$\hat{\rho} \approx 0$	$DW \approx 2$
Correlación positiva	$1 > \rho > 0$	$1 > \hat{\rho} > 0$	$DW < 2$
Correlación negativa	$-1 < \rho < 0$	$-1 < \hat{\rho} < 0$	$DW > 2$

Naturalmente la regla que se presenta en el recuadro 2 es únicamente intuitiva. Para tener una decisión con mayor grado de certidumbre se deberá efectuar una prueba de hipótesis.

El estadístico DW nos permite contrastar tres diferentes hipótesis nulas, como se reportan en el recuadro 6.3.

Recuadro 6.3 Estadístico DW en casos de Autocorrelación

$H_0 : \rho = 0$	$\rho = 0$	$H_A : \rho \neq 0$
H_o: No autocorrelación Positiva	$1 > \rho > 0$	$H_A : \rho > 0$
H_o: No autocorrelación Negativa	$-1 < \rho < 0$	$H_A : \rho < 0$

[5] Una manera alternativa para comprobar la hipótesis nula, cuando se tienen más de 20 observaciones de un mismo signo es emplear como estadístico de prueba $RA = \frac{k - E(k)}{\sqrt{Var(k)}}$. La hipótesis nula puede ser rechazada si $|RA| > z_{\frac{\alpha}{2}}$.

Durbin y Watson (1951) encontraron la distribución de su estadístico DW y la tabularon (Valores que se reportan al final de este capítulo). Esta distribución implica dos valores críticos: el límite inferior (d_l) y el límite superior (d_u).

Es importante tener en cuenta que si el estadístico de prueba se encuentra entre $d_l < DW < d_u$ o entre $4 - d_u < DW < 4 - d_l$, entonces la prueba no puede darnos una decisión y por lo tanto no sería posible emplearla. En el recuadro 6.4 se presentan las diferentes hipótesis nulas disponibles y cómo se toma la decisión para cada caso.

Recuadro 6.4 Estadístico DW en casos de Autocorrelación

$\mathbf{H_0}$	**Regla de decisión**	**Decisión**
$H_0 : \rho = 0$	$d_u < DW < 4 - d_u$	No rechazar H_0
H_0: No autocorrelación positiva	$0 < DW < d_l$	Rechazar H_0
H_0: No autocorrelación Negativa	$4 - d_l < DW < 4$	Rechazar H_0

Sobre esta prueba es importante destacar varios aspectos:

- El DW no tiene sentido si no hay intercepto (ver Durbin y Watson (1951)).
- Los valores d_u y d_l dependen del número de observaciones y del número de regresores.
- El DW depende del supuesto que las X's sean no estocásticas (ver Durbin y Watson (1951)).
- La prueba sólo tiene poder contra procesos AR(1).
- Esta prueba tampoco aplica en los casos en que existen variables dependientes rezagadas en la derecha del modelo; por ejemplo, para el modelo $Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + \varepsilon_t$ este estadístico no aplica.

6.2.3. Prueba h de Durbin

Como se mencionó anteriormente, si el modelo emplea la variable dependiente rezagada como explicativa, la prueba de Durbin-Watson no aplica. Para solucionar este problema Durbin (1970) sugirió el siguiente estadístico:

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n\left(\widehat{Var\left(\hat{\alpha}\right)}\right)}}$$

donde:

$$Y_t = \beta_0 + \beta_1 X_{1i} + ... + \beta_k X_{ki} + \alpha Y_{t-1} + \varepsilon_t$$

Por lo tanto, es relativamente sencillo demostrar que:

$$h = \left(1 - \frac{DW}{2}\right) \sqrt{\frac{n}{1 - n\left(\widehat{Var\left(\hat{\alpha}\right)}\right)}} \tag{6.5}$$

Durbin (1970) demostró que este estadístico de prueba sigue una distribución estándar normal ($h \sim N(0,1)$) y por lo tanto, si se cumple que $|h| > z_{\frac{\alpha}{2}}$ entonces es posible rechazar $H_0 : \rho = 0$ a favor de la $H_A : \rho \neq 0$. Finalmente, como podemos apreciar en la ecuación 6.5, esta prueba no es válida en los casos en que $n\left(\widehat{Var(\hat{\alpha})}\right) \geq 1$.

6.2.4. Prueba de Box-Pierce y Ljung-Box

Otra aproximación para comprobar la existencia o no de autocorrelación es determinar si las autocorrelaciones a diferentes rezagos son o no iguales a cero. Box y Pierce (1970) diseñan una prueba basada en la autocorrelación muestral de los errores que permite detectar la existencia de errores con procesos más persistentes que AR(1). Recordemos que la autocorrelación poblacional se define de la siguiente forma:

$$\gamma_j = \frac{Cov(\varepsilon_t, \varepsilon_{t-j})}{\sqrt{Var(\varepsilon_t)\, Var(\varepsilon_{t-j})}} = \frac{Cov(\varepsilon_t, \varepsilon_{t-j})}{\sigma_\varepsilon^2}$$

Y la correspondiente autocorrelación muestral es:

$$\hat{\gamma}_j = \frac{\sum_{t=k+1}^{n} (\hat{\varepsilon}_t - \bar{\varepsilon})(\hat{\varepsilon}_{t-j} - \bar{\varepsilon})}{\sum_{t=1}^{n} (\hat{\varepsilon}_t - \bar{\varepsilon})^2} = \frac{\sum_{t=j+1}^{n} \hat{\varepsilon}_t \hat{\varepsilon}_{t-j}}{\sum_{t=1}^{n} (\hat{\varepsilon}_t)^2}$$

Box y Pierce (1970) definen una prueba que permite determinar si las primeras s autocorrelaciones son conjuntamente iguales a cero o no. Es decir, permite comprobar la hipótesis nula de un error no autocorrelacionado (las correlaciones a los s rezagos son cero), versus la hipótesis alterna de la existencia de algún tipo de autocorrelación (por lo menos una autocorrelación no es cero). Para comprobar esta hipótesis nula, Box y Pierce (1970) sugieren el estadístico Q:

$$Q = n \sum_{j=1}^{s} r_k^2 \sim_a \chi_s^2$$

donde s corresponden al número de rezagos que se desean considerar dentro de la prueba. Ellos demuestran que su estadístico sigue una distribución Chi-cuadrado con s grados de libertad (χ_s^2). Por lo tanto, será posible rechazar la H_0 (error no autocorrelacionado) si se cumple que $Q > \chi_s^2$.

Sin embargo, la prueba de Box-Pierce sólo es válida para muestras grandes ($n > 20$), para resolver este inconveniente Ljung y Box (1979) proponen una modificación del estadístico anterior para que presente un mejor comportamiento en muestras pequeñas. El estadístico de Ljung-Box corresponde a:

$$Q' = n(n+2) \sum_{k=1}^{s} \frac{r_j^2}{n+j}$$

Este estadístico funciona y posee la misma distribución que el de la prueba de Box-Pierce.

Finalmente, es importante mencionar que una práctica muy común es realizar esta prueba para un número relativamente grande de rezagos. Es decir, hacer las correspondientes pruebas para diferentes rezagos; por ejemplo, se calculan los correspondientes estadísticos para comprobar las siguientes

hipótesis alternas:

$$\begin{aligned} &H_0 : \gamma_1 = 0 \\ &H_0 : \gamma_1 = \gamma_2 = 0 \\ &\vdots \\ &H_0 : \gamma_1 = \gamma_2 = \ldots \gamma_{n/3} = 0 \end{aligned}$$

La decisión de si los errores están o no autocorrelacionados se toma teniendo en cuenta las decisiones de cada una de estas pruebas.

6.2.5. Prueba de Breusch-Godfrey

Breusch (1978); Godfrey (1978) diseñaron una prueba que permite comprobar la hipótesis nula de no autocorrelación versus la alterna de que el error sigue un proceso auto-regresivo de orden p. Esta prueba también es conocida como la prueba del multiplicador de Lagrange o prueba LM (por su sigla en inglés: Lagrange multiplier test).

Esta prueba se basa en una idea muy sencilla. Si existe autocorrelación en los errores, entonces éstos son explicados por sus valores pasados, pero si no hay autocorrelación entonces los valores pasados de los errores no pueden explicar el comportamiento actual del error.

Así, para probar la hipótesis nula de no autocorrelación versus la alterna de unos errores con un proceso $AR(s)$, se pueden emplear los residuos estimados de la regresión bajo estudio para comprobar si los valores pasados del error sirven o no para explicar el error del periodo t. Es decir, la prueba LM implica los siguientes pasos:

1. Estime el modelo de regresión original:

$$y_t = \beta_1 + \beta_2 X_{2,t} + \beta_3 X_{3,t} + \ldots + \beta_k X_{k,t} + \varepsilon_t$$

   y obtenga la serie de los errores estimados ($\hat{\varepsilon}_t$).

2. Estime la siguiente regresión auxiliar:

$$\hat{\varepsilon}_t = \alpha_1 + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \ldots + \alpha_k X_{k,t} + \omega_1 \hat{\varepsilon}_{t-1} + \omega_2 \hat{\varepsilon}_{t-2} + \ldots + \omega_s \hat{\varepsilon}_{t-s} + \xi_t$$

3. Empleando el R^2 de la regresión auxiliar calcule el estadístico LM de la siguiente manera:[6]

$$LM = (n - s) \times R^2$$

4. Compare el estadístico LM con el valor crítico de la distribución Chi-cuadrado con s grados de libertad. Se rechazará la hipótesis nula si $LM > \chi^2_{s,\alpha}$.

Al igual que la prueba de Box-Pierce, cuando se emplea esta prueba normalmente se realizan las pruebas para diferentes hipótesis nulas y se toma la decisión basándose en el conjunto de los resultados.

[6] Algunos paquetes estadísticos calculan el estadístico LM multiplicando el R^2 por n y no por $(n - s)$. Si el tamaño de la muestra es grande, estas dos aproximaciones son equivalentes, en caso contrario es mejor multiplicar por $(n - s)$.

6.3. Solución a la autocorrelación

6.3.1. Solución por diferencias generalizadas

Si conocemos la naturaleza de la autocorrelación entonces podemos usar una transformación de la muestra del modelo original para construir una muestra sin este problema. Este método se conoce como transformación de diferencias generalizadas.

Es decir, partiendo del modelo 6.5 con errores $AR(1)$, se le puede restar el mismo modelo 6.5 rezagado un período y multiplicado por la correlación. De tal manera que se obtiene:

$$\begin{aligned} Y_t - \rho Y_{t-1} =& \beta_1 (1-\rho) + \beta_2 (X_{2t} - \rho X_{2t-1}) + \beta_3 (X_{3t} - \rho X_{3t-1}) + \dots \\ &+ \beta_k (X_{kt} - \rho X_{kt-1}) + \varepsilon_t - \rho \varepsilon_{t-1} \end{aligned}$$

Reparametrizando tenemos:

$$Y_t^* = \beta_1 (1-\rho) + \beta_2 X_{2t}^* + \beta_3 X_{3t}^* + \dots + \beta_k X_{kt}^* + \nu_t$$

donde

$$Y_t^* = Y_t - \rho Y_{t-1}, \quad X_{2t}^* = X_{2t} - \rho X_{2t-1} \quad \text{y} \quad \mathrm{X}_{3\mathrm{t}}^* = \mathrm{X}_{3\mathrm{t}} - \rho \mathrm{X}_{3\mathrm{t}-1} \tag{6.6}$$

Así, el modelo transformado (6.6) ya no tiene problemas de autocorrelación. Sin embargo, en la realidad es prácticamente imposible que conozcamos la naturaleza del problema de autocorrelación con certeza. Durbin (1960) plantea la solución que se discute en la siguiente sección.

6.3.2. Método de Durbin

El método de Durbin (1960) permite implementar el método de diferencias generalizadas al estimar en un primer paso el valor de ρ.[7] Este método implica los siguientes tres pasos:

1. Corra la regresión de la variable dependiente en función de las variables explicativas del modelo original, además incluya las mismas variables independientes rezagadas un período y la variable dependiente rezagada un período. Es decir:

$$\begin{aligned} Y_t =& \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \beta_{k+1} X_{2t-1} \\ &+ \beta_{k+2} X_{3t-1} + \dots + \beta_{k+(k-1)} X_{kt-1} + \rho Y_{t-1} + \varepsilon_t \end{aligned}$$

2. De la anterior estimación se obtiene el valor estimado de ρ. A partir de ese valor estimado, realice las siguientes transformaciones:

$$Y_t^* = Y_t - \hat{\rho} Y_{t-1}, \quad X_{2t}^* = X_{2t} - \hat{\rho} X_{2t-1} \quad \text{y} \quad \mathrm{X}_{3\mathrm{t}}^* = \mathrm{X}_{3\mathrm{t}} - \hat{\rho} \mathrm{X}_{3\mathrm{t}-1}$$

3. Finalmente estime el siguiente modelo:

$$Y_t^* = \beta_1^* + \beta_2 X_{2t}^* + \beta_3 X_{3t}^* + \dots + \beta_k X_{kt}^* + \nu_t$$

   Donde $\beta_1^* = \beta_1 (1 - \hat{\rho})$

[7] Este método corresponde al método de mínimos cuadrados factibles

6.4. Sostenibilidad de la política fiscal en Colombia

Durante años se ha argumentado que en Colombia la política fiscal ha sido manejada de forma tal que se ha convertido en algo insostenible, además existen constantes debates que cuestionan el uso de estas políticas con el fin de incidir en la economía en el largo plazo. En este ejercicio se empleará un análisis de series de tiempo en el periodo 1964 a 2004, con el cual será posible determinar si la política fiscal es sostenible o no. Además, aplicaremos las diferentes pruebas de detección de la autocorrelación que aprendimos en este capítulo y de ser el caso aplicaremos la solución a este problema, claro si es que está presente

Basados en previos estudios realizados por Bohn (1995) en los que se analiza la estabilidad de la política fiscal, este documento examina la estabilidad en el largo plazo de estas políticas mediante la respuesta que tiene el superávit primario (como porcentaje del PIB) ante cambios en la deuda pública (como porcentaje del PIB). El razonamiento que se encuentra detrás de esta metodología es que la política fiscal es estable ya que los incrementos en la deuda generan un aumento en los ingresos mayor a los gastos y además dicho endeudamiento amortiza la deuda mediante el pago de intereses, de esta forma la política fiscal y el endeudamiento no divergen impidiendo un crecimiento no sostenible que finalmente estallaría como si fuese un efecto burbuja.

Para probar la sostenibilidad fiscal en el caso colombiano se usará el enfoque utilizado por Greiner et al. (2005) quienes lo aplicaron a la economía alemana.[8]

La especificación del modelo es la siguiente:

$$SR_t = \beta_0 + \delta DGR_{t-1} + \beta_1 Z_t + \varepsilon_t \tag{6.7}$$

Donde SR_t y DGR_{t-1} representan el déficit primario y la deuda pública ambas como porcentaje del PIB respectivamente, Z_t es un vector que incluye a las otras variables que afectan el superávit primario, y ε_t es el término de error el cual cumple los supuestos del teorema de Gauss-Markov. En nuestro caso, el modelo a estimar será:

$$SR_t = \beta_0 + \delta DGR_{t-1} + \alpha_1 r_t + \alpha_2 BC_t + \varepsilon_t \tag{6.8}$$

Siendo r_t la tasa de interés bancaria pasiva a 90 días y BC_t corresponde al ciclo del PIB real expresado en millones de pesos constantes del año 1994 filtrado dos veces.[9] Según Bohn (1995) la condición de estabilidad implica que para valores de $\delta \geq 0$ la política fiscal es sostenible y se mantiene la restricción intertemporal del presupuesto y para valores de $\delta < 0$ diverge y por lo tanto es insostenible.

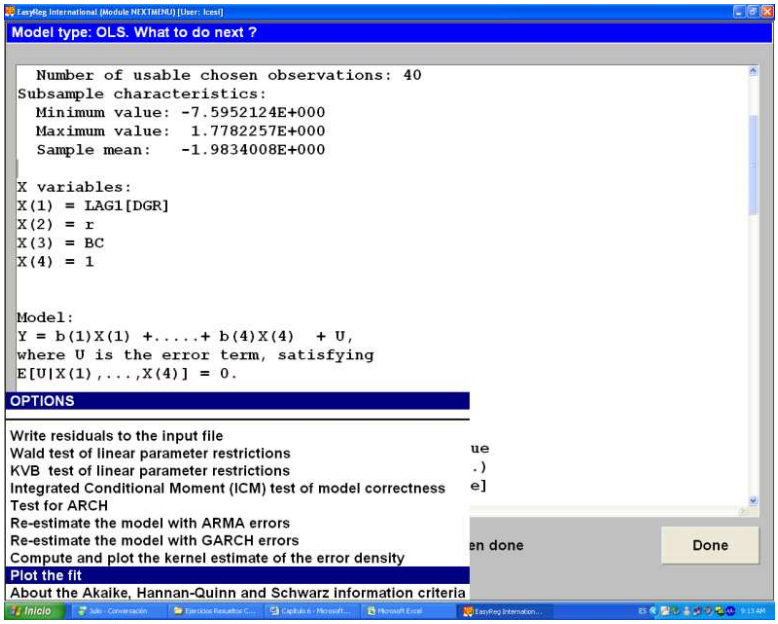
Los datos que emplearemos son anuales y comprenden el periodo entre 1964 y 2004, estos se encuentran en el archivo eauto.xls.

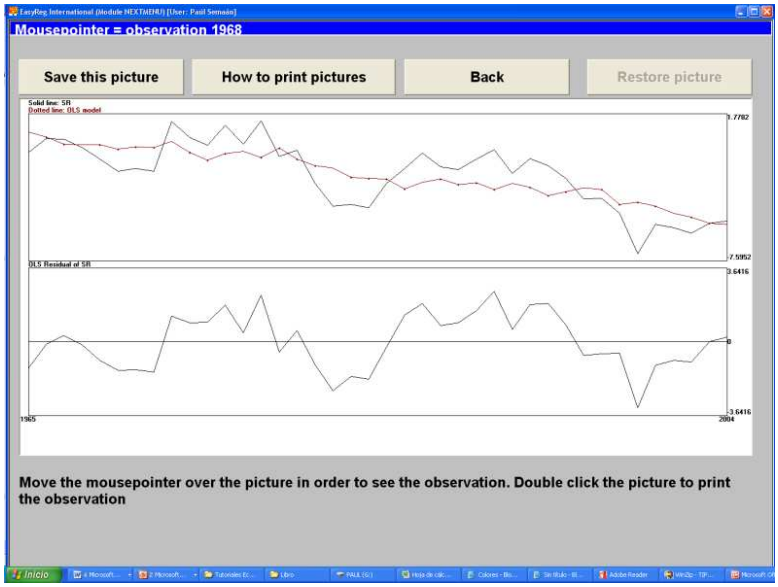
6.4.1. Análisis gráfico de los residuos y prueba de Durbin-Watson

Cargue los datos y corra la regresión correspondiente al modelo (6.9) empleando el método de mínimos cuadrados ordinarios (Para ver detalles de cómo hacer esto revise el Capítulo 1, no olvide rezagar en EasyReg la variable DGR_t). Haga clic en el botón "Options". Note que cuando trabaja con series de tiempo, existe una posibilidad más en el anterior menú que no hay cuando se trabaja con datos de corte transversal. Para gráficas los errores, haga clic en la opción "Plot the fit".

[8] Para otra aproximación al problema ver Alonso et al. (1997).

[9] Este procedimiento se realiza con el fin de extraer del PIB el comportamiento cíclico. Aunque es un procedimiento relativamente sencillo, de todas maneras no haremos énfasis en esto





En el panel superior, EasyReg grafica los valores estimados de la variable dependiente ($\widehat{SR}_t$) y sus correspondientes valores observados SR_t. En el panel inferior, se encuentra el gráfico de los residuos estimados ($\hat{\varepsilon}_t$) versus el tiempo.[10] En este gráfico se ven claros síntomas de autocorrelación positiva, pues los residuos parecen tener una fuerte persistencia en su signo. En otras palabras, cuando la serie alcanza valores positivos, estos tienden a persistir, lo mismo ocurre con valores negativos.

Ahora haga clic en el botón "Back", esto lo llevará de retorno a los resultados del modelo estimado. En los resultados, inmediatamente debajo del estadístico F global, observará el estadístico Durbin-Watson. En este caso, tenemos que DW es igual a 0.808531. Como lo discutimos, un DW menor que dos puede implicar la presencia de autocorrelación. Para efectuar una prueba formal de la presencia o no de autocorrelación, necesitamos consultar la tabla de valores críticos para el estadístico Durbin-

[10]Recuerde que usted puede grabar este gráfico haciendo clic en la opción "Save this picture".

Watson. En este caso, con un nivel de significancia del 5 % (3 variables explicatorias y 40 observaciones) tenemos que d_l=1.34 y d_u= 1.66.

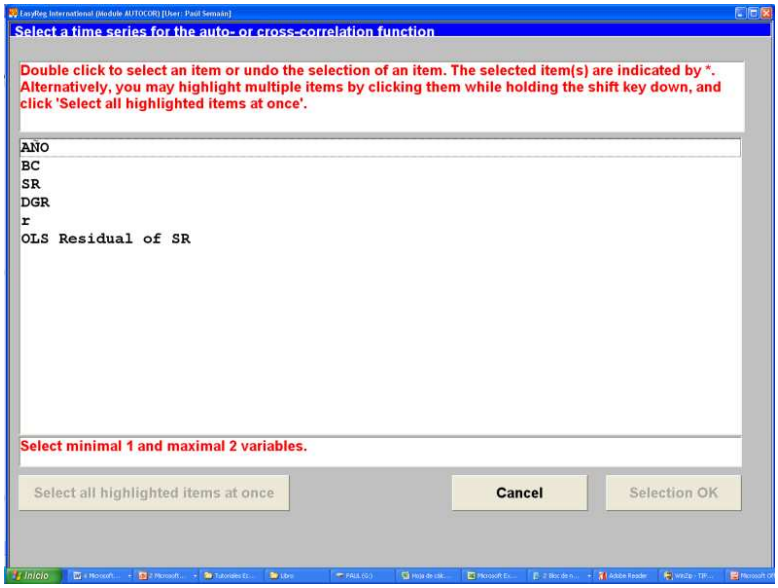
Noten que el DW no está entre d_l y d_u ni entre $4-d_u$ y $4-d_l$ por lo tanto la prueba aplica. Además tenemos que $0 < DW < d_l$, de esta forma, podemos rechazar la hipótesis nula de no autocorrelación positiva a favor de la hipótesis alterna (autocorrelación positiva).

6.4.2. Cálculo de la autocorrelación para los residuos, prueba de Box-Pierce y de Ljung-Box y gráfico de las autocorrelaciones

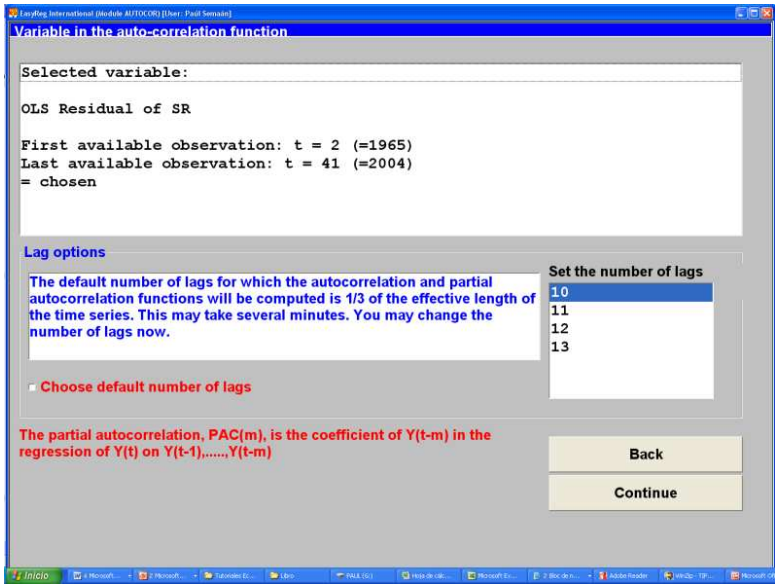
Haga clic en el menú "Options". y seleccione la opción "Write residuals to the input file", esto creará una nueva variable con el nombre "OLS Residuals of SR". Ahora regrese a la ventana principal de EasyReg (Haga clic en la opción "Done").

Ahora, haga clic en el botón "Menu / Data analysis / Auto/Cross correlations".

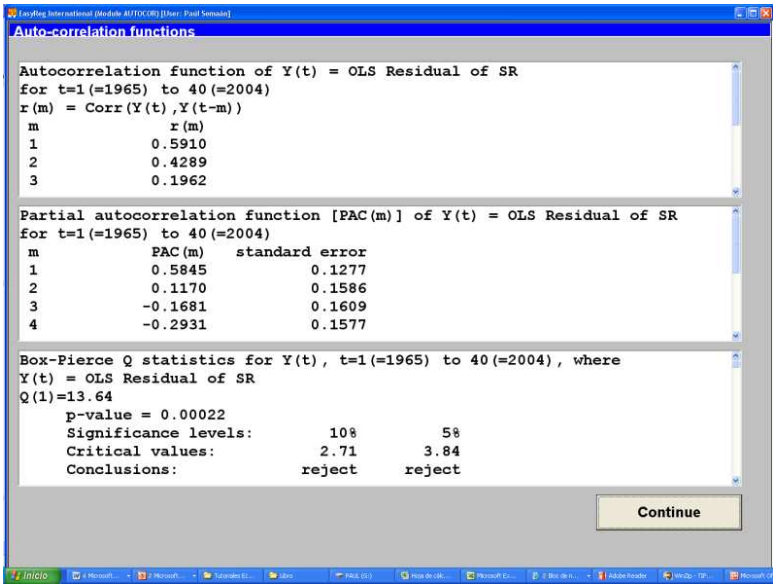
Verá la siguiente ventana en la que debe escoger la variable para la cual desea conocer las autocorrelaciones y correlaciones parciales, así como los correspondientes tests de Box-Pierce y de Ljung-Box.



Escoja la variable "OLS Residual of SR" y haga clic en el botón "Selection OK". Haga clic en el botón "No" de la siguiente ventana y posteriormente en "Continue". Observará la siguiente ventana.



En esta ventana usted puede escoger el número de rezagos para los cuales quiere calcular la autocorrelación y la correlación parcial. El número de rezagos pre-establecido es un tercio de las observaciones. En nuestro caso, esto implicaría calcular hasta la autocorrelación 13, esto es excesivo. En la ventana "Set the number of lags", escoja el mínimo número de rezagos, es decir, 10 en este caso. Haga clic en el botón "Continue". Observará la siguiente ventana:



En el primer panel, observará la autocorrelación para la serie de los residuos. En el segundo panel, encontrará la autocorrelación parcial y su respectiva desviación estándar. En el tercer panel, hallará los estadísticos de Box-Pierce y de Ljung-Box. Los resultados de ambas pruebas están reportados en los recuadros 6.5 y 6.6.

Recuadro 6.5 Resultados del Test de Box-Pierce

Box-Pierce Q statistics for Y(t), t=1(=1965) to 40(=2004), where
Y(t) = OLS Residual of SR

Q(1)=13.64		
p-value = 0.00022		
Significance levels:	10 %	5 %
Critical values:	2.71	3.84
Conclusions:	reject	reject
Q(2)=20.83		
p-value = 0.00003		
Significance levels:	10 %	5 %
Critical values:	4.61	5.99
Conclusions:	reject	reject
Q(3)=22.31		
p-value = 0.00006		
Significance levels:	10 %	5 %
Critical values:	6.25	7.81
Conclusions:	reject	reject
Q(4)=22.57		
p-value = 0.00015		
Significance levels:	10 %	5 %
Critical values:	7.78	9.49
Conclusions:	reject	reject
Q(5)=24.91		
p-value = 0.00014		
Significance levels:	10 %	5 %
Critical values:	9.24	11.07
Conclusions:	reject	reject

Q(6)=31.12		
p-value = 0.00002		
Significance levels:	10 %	5 %
Critical values:	10.64	12.59
Conclusions:	reject	reject
Q(7)=35.18		
p-value = 0.00001		
Significance levels:	10 %	5 %
Critical values:	12.02	14.07
Conclusions:	reject	reject
Q(8)=43.27		
p-value = 0.00000		
Significance levels:	10 %	5 %
Critical values:	13.36	15.51
Conclusions:	reject	reject
Q(9)=48.80		
p-value = 0.00000		
Significance levels:	10 %	5 %
Critical values:	14.68	16.92
Conclusions:	reject	reject
Q(10)=50.81		
p-value = 0.00000		
Significance levels:	10 %	5 %
Critical values:	15.99	18.31
Conclusions:	reject	reject

Los resultados tabulados en forma organizada se reportan en el cuadro 6.1. Note que para probar la H_0: (no autocorrelación) versus H_A: (por lo menos una autocorrelación es diferente de cero), podemos considerar distintos rezagos. Siempre existirá el problema de determinar cuántos rezagos emplear para el test. La prueba puede no detectar la autocorrelación con rezagos de orden superior, si se consideran pocos rezagos. Sin embargo, si se emplean muchos rezagos, la prueba puede tener bajo poder. Para más detalles, consulte Ljung y Box (1979).

Como se puede observar en los recuadros 6.5 y 6.6 los estadísticos de Box-Pierce y los de Ljung-Box son iguales. En este caso, estos dos estadísticos son iguales dado que el número de observaciones es relativamente grande y el número de regresores es pequeño.[11] Este estadístico se compara con el valor de la distribución Chi-cuadrado con un grado de libertad (2.71 y 3.84 para niveles de significancia de 10 % y 5 %, respectivamente). Entonces se puede rechazar la hipótesis nula a favor de la presencia de un error que sigue un proceso autorregresivo. Noten que si se consideran más rezagos, la decisión sigue siendo la misma.

[11]Note que los estadísticos de Box-Pierce y Ljung-Box son $Q = n \sum_{k=1}^{p} r_k^2$ y $Q' = n\,(n+2) \sum_{k=1}^{p} \frac{r_k^2}{n+k}$, respectivamente. Así cuando n es lo suficientemente grande y el número de regresores es pequeño, entonces $Q' = n \sum_{k=1}^{p} \frac{(n+2)r_k^2}{n+k} \approx n \sum_{k=1}^{p} r_k^2 = Q$.

Recuadro 6.6 Resultados del Test de Ljung-Box

Box-Pierce Q statistics for Y(t), t=1(=1965) to 40(=2004), where
Y(t) = OLS Residual of SR

Q(1)=13.64		
p-value = 0.00022		
Significance levels:	10 %	5 %
Critical values:	2.71	3.84
Conclusions:	reject	reject
Q(2)=20.83		
p-value = 0.00003		
Significance levels:	10 %	5 %
Critical values:	4.61	5.99
Conclusions:	reject	reject
Q(3)=22.31		
p-value = 0.00006		
Significance levels:	10 %	5 %
Critical values:	6.25	7.81
Conclusions:	reject	reject
Q(4)=22.57		
p-value = 0.00015		
Significance levels:	10 %	5 %
Critical values:	7.78	9.49
Conclusions:	reject	reject
Q(5)=24.91		
p-value = 0.00014		
Significance levels:	10 %	5 %
Critical values:	9.24	11.07
Conclusions:	reject	reject
Q(6)=31.12		
p-value = 0.00002		
Significance levels:	10 %	5 %
Critical values:	10.64	12.59
Conclusions:	reject	reject
Q(7)=35.18		
p-value = 0.00001		
Significance levels:	10 %	5 %
Critical values:	12.02	14.07
Conclusions:	reject	reject
Q(8)=43.27		
p-value = 0.00000		
Significance levels:	10 %	5 %
Critical values:	13.36	15.51
Conclusions:	reject	reject
Q(9)=48.80		
p-value = 0.00000		
Significance levels:	10 %	5 %
Critical values:	14.68	16.92
Conclusions:	reject	reject
Q(10)=50.81		
p-value = 0.00000		
Significance levels:	10 %	5 %
Critical values:	15.99	18.31
Conclusions:	reject	reject

Cuadro 6.1: Resultados de las pruebas de Box-Pierce y Ljung-Box

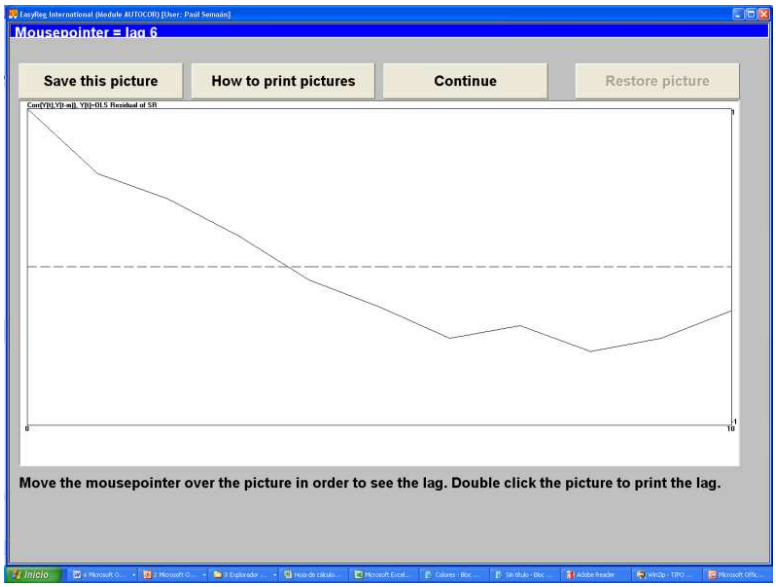
	Estadístico	
Rezago	Box-Pierce	Ljung-Box
1	13.64***	13.64***
2	20.83***	20.83***
3	22.31***	22.31***
4	22.57***	22.57***
5	24.91***	24.91***
6	31.12***	31.12***
7	35.18***	35.18***
8	43.27***	43.27***
9	48.8***	48.8***
10	50.81***	50.81***

(*) Rechaza la H_0 con un nivel de significancia: 10 %

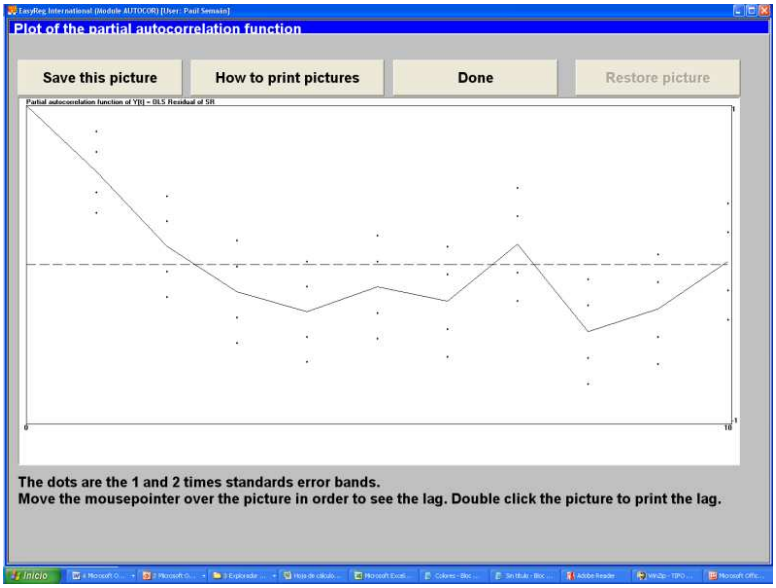
(**) Rechaza la H_0 con un nivel de significancia: 5 %

(***) Rechaza la H_0 con un nivel de significancia: 1 %

Haga clic en el botón "Continue". Enseguida observará el gráfico de las correlaciones estimadas versus el número de rezagos del residuo.



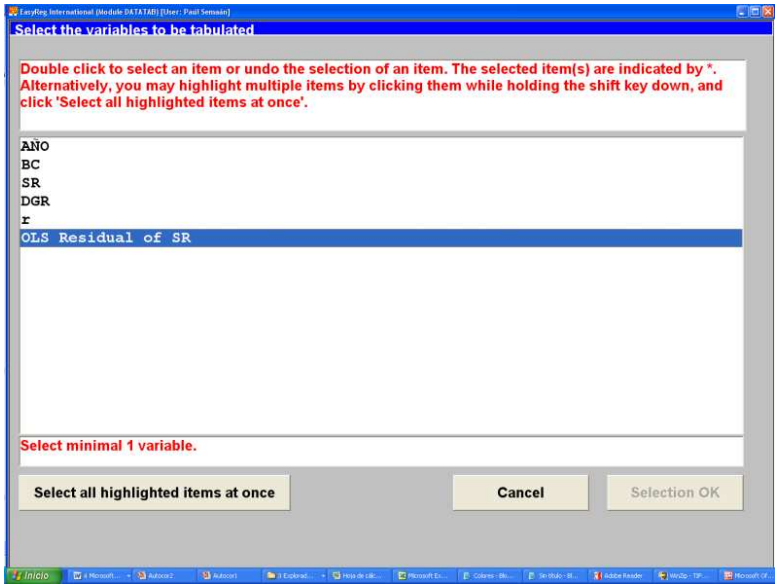
Si hace clic en el botón "Continue" observará el gráfico de la función de autocorrelación parcial.



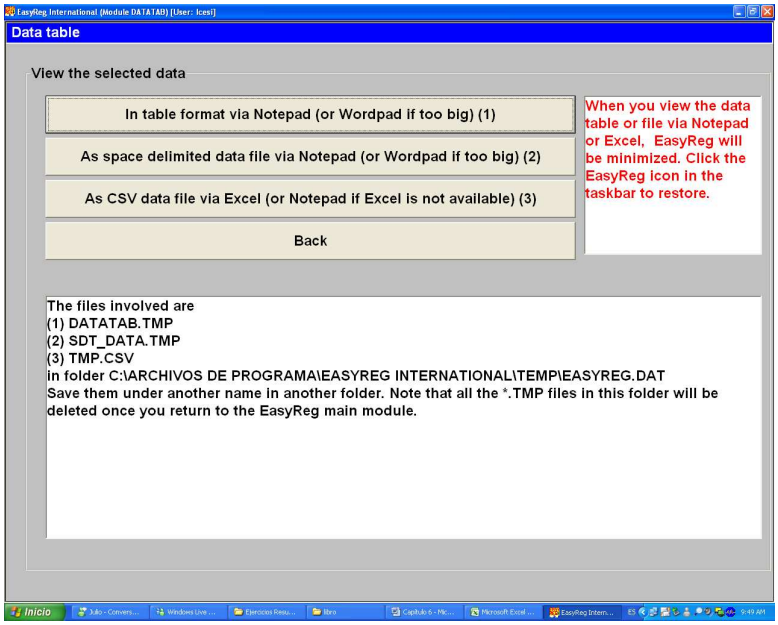
6.4.3. Prueba de Rachas

Para efectuar esta prueba necesitamos la serie de los residuos. En el menú principal elija "Menu", "Data analisis" y luego "Make data table"

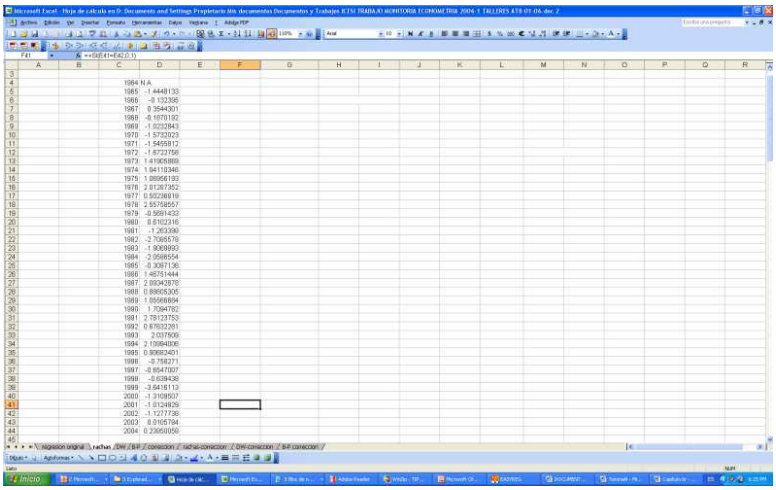
En la siguiente ventana elija la variable "OLS residual of SR"



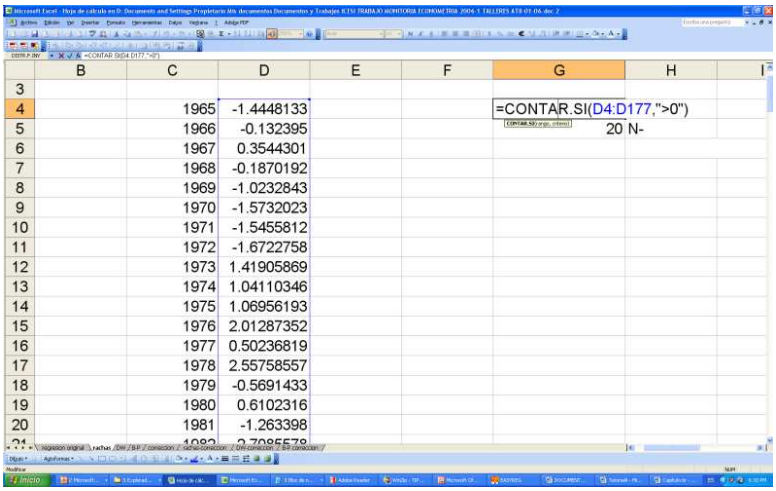
En la siguiente ventana elija la opción "In table format via Notepad (or Wordpad if too big)"



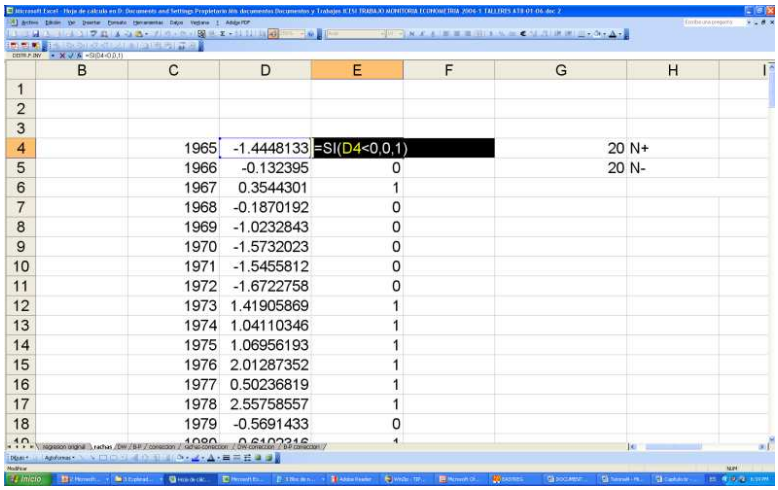
Seleccione todos los datos y luego péguelos en un archivo de Excel (si tiene problemas con el pegado correcto de los datos remítase al capítulo 4). Ahora tenemos lo siguiente:



Con el condicional CONTAR.SI(D4:D177,">0") podemos calcular el valor N_+ y cambiando el signo podemos obtener N_-. Coincidencialmente en este caso N_+=20 y N_-=20

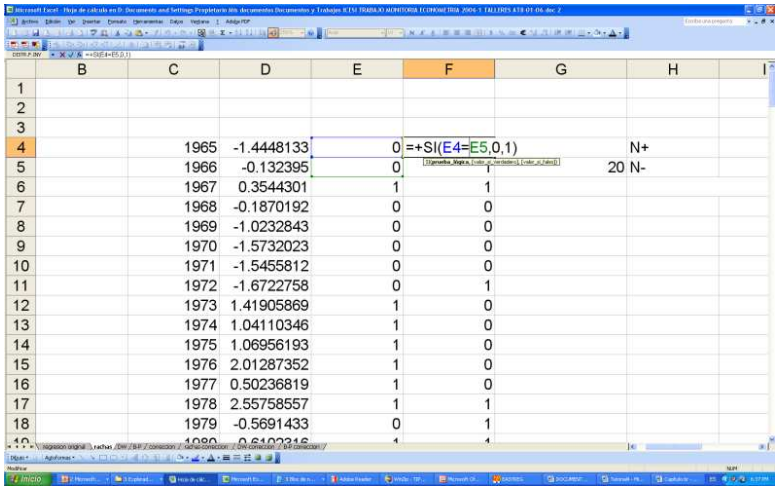


Ahora en la columna siguiente a la que se encuentran los residuos, creemos una que tenga el valor de 0 cuando el residuo sea negativo y uno cuando sí es positivo. En este caso lo podemos hacer con el condicional SI(D4<0,0,1)



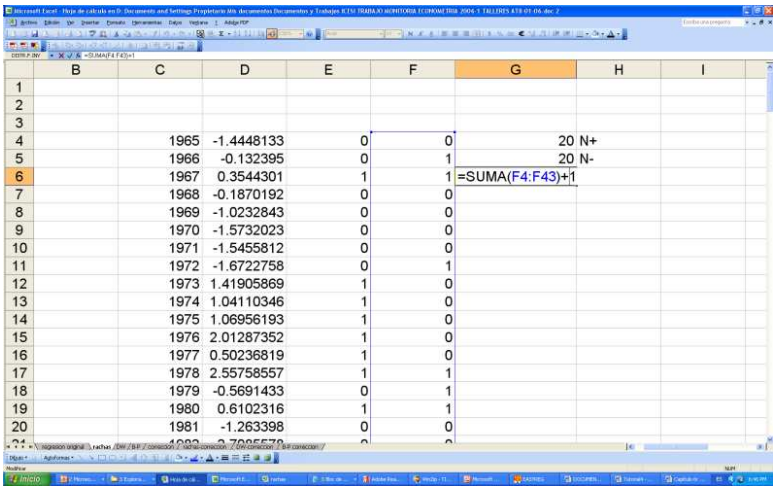
	B	C	D	E	F	G	H	I
1								
2								
3								
4		1965	-1.4448133	=SI(D4<0,0,1)		20	N+	
5		1966	-0.132395	0		20	N-	
6		1967	0.3544301	1				
7		1968	-0.1870192	0				
8		1969	-1.0232843	0				
9		1970	-1.5732023	0				
10		1971	-1.5455812	0				
11		1972	-1.6722758	0				
12		1973	1.41905869	1				
13		1974	1.04110346	1				
14		1975	1.06956193	1				
15		1976	2.01287352	1				
16		1977	0.50236819	1				
17		1978	2.55758557	1				
18		1979	-0.5691433	0				

Ahora podemos contar el número de cambios de signo, para esto ubiquémonos en la siguiente columna y en este caso utilicemos el condicional SI(E4=E5,0,1)



	B	C	D	E	F	G	H	I
1								
2								
3								
4		1965	-1.4448133	0	=+SI(E4=E5,0,1)		N+	
5		1966	-0.132395	0		20	N-	
6		1967	0.3544301	1	1			
7		1968	-0.1870192	0	0			
8		1969	-1.0232843	0	0			
9		1970	-1.5732023	0	0			
10		1971	-1.5455812	0	0			
11		1972	-1.6722758	0	1			
12		1973	1.41905869	1	0			
13		1974	1.04110346	1	0			
14		1975	1.06956193	1	0			
15		1976	2.01287352	1	0			
16		1977	0.50236819	1	0			
17		1978	2.55758557	1	1			
18		1979	-0.5691433	0	1			

Finalmente tenemos que las Rachas (k) serán iguales al número de cambios de signo más 1. En este caso tenemos un total de 11 Rachas.



	B	C	D	E	F	G	H	I
1								
2								
3								
4		1965	-1.4448133	0	0	20 N+		
5		1966	-0.132395	0	1	20 N-		
6		1967	0.3544301	1	1	=SUMA(F4:F43)+1		
7		1968	-0.1870192	0	0			
8		1969	-1.0232843	0	0			
9		1970	-1.5732023	0	0			
10		1971	-1.5455812	0	0			
11		1972	-1.6722758	0	1			
12		1973	1.41905869	1	0			
13		1974	1.04110346	1	0			
14		1975	1.06956193	1	0			
15		1976	2.01287352	1	0			
16		1977	0.50236819	1	0			
17		1978	2.55758557	1	1			
18		1979	-0.5691433	0	1			
19		1980	0.6102316	1	1			
20		1981	-1.263398	0	0			

Además, con valores de N_+ y N_- ambos iguales a 20, encontrar el valor esperado y la varianza de (k)

$$E(k) = \frac{2N_+N_-}{N_+ + N_-} + 1 = 21$$

$$Var(k) = \frac{2N_+N_-\left(2N_+N_- - N_+ - N_-\right)}{\left(N_+ + N_-\right)^2\left(N_+ + N_- - 1\right)} = 9{,}74$$

Entonces tenemos que

$$RA = \frac{k - E(k)}{\sqrt{Var(k)}} = -3{,}2$$

Recordemos que rechazamos H_0 si $|RA| > z_{\frac{\alpha}{2}}$, en este caso tenemos que a un nivel de significancia del $1\,\%$, $z_{\frac{\alpha}{2}} = 2{,}575$ y por lo tanto rechazamos la hipótesis nula de no autocorrelación.

6.4.4. Prueba de Breusch-Godfrey

Para realizar esta prueba debemos estimar el modelo 6.7 y crear la serie de los residuos, lo cual ya hemos hecho. Posteriormente estimamos la siguiente regresión auxiliar para cada una de las hipótesis alternas que queramos probar; de manera general tenemos:

$$\hat{\varepsilon}_t = \alpha_1 + \alpha_2 X_{2,t} + \alpha_3 X_{3,t} + \ldots + \alpha_k X_{k,t} + \omega_1\hat{\varepsilon}_{t-1} + \omega_2\hat{\varepsilon}_{t-2} + \ldots + \omega_s\hat{\varepsilon}_{t-s} + \xi_t$$

Es decir, para contrastar que los errores de la regresión obedecen a un proceso $AR(1)$ estimamos:

$$\hat{\varepsilon}_t = \beta_0 + \delta DGR_{t-1} + \alpha_1 r_t + \alpha_2 BC_t + \omega_1\hat{\varepsilon}_{t-1} + \xi$$

Para un proceso $AR(2)$:

$$\hat{\varepsilon}_t = \beta_0 + \delta DGR_{t-1} + \alpha_1 r_t + \alpha_2 BC_t + \omega_1\hat{\varepsilon}_{t-1} + \omega_2\hat{\varepsilon}_{t-2} + \xi$$

Y así sucesivamente

De cada una de las anteriores estimaciones utilizamos el R^2 para calcular el estadístico LM de la siguiente manera:

$$LM = (n - s) \times R^2$$

Posteriormente comparamos el estadístico LM con el valor crítico de la distribución Chi-cuadrado con s grados de libertad. Rechazaremos la hipótesis nula sí $LM > \chi^2_{s,\alpha}$. Estos resultados se resumen en el cuadro 6.2:

Cuadro 6.2: Resultados de las pruebas de Breusch-Godfrey

	Proceso Autorregresivo del Error				
	$AR(1)$	$AR(2)$	$AR(3)$	$AR(4)$	$AR(5)$
R^2	0.3137	0.4364	0.438	0.4619	0.4856
LM	11.9206***	15.7104***	14.892***	14.7808**	14.568**

(*) Rechaza la H_0 con un nivel de significancia: 10 %

(**) Rechaza la H_0 con un nivel de significancia: 5 %

(***) Rechaza la H_0 con un nivel de significancia: 1 %

En resumen, las tres pruebas muestran evidencia a favor de la existencia de autocorrelación en los residuos estimados.

6.4.5. Método de corrección de Durbin

Recuerde que los pasos para efectuar esta corrección son los siguientes:

1. Estime el modelo:

$$SR_t = \beta_0 + \beta_1 DGR_{t-1} + \beta_2 r_t + \beta_3 BC_t + \beta_4 DGR_{t-2} + \beta_5 r_{t-1} + \beta_6 BC_{t-1} + \rho SR_{t-1} + \varepsilon_t$$

2. Realice las siguientes transformaciones

$$SR^*_t = SR_t - \hat{\rho} SR_{t-1}, \quad DGR^*_{t-1} = DGR_{t-1} - \hat{\rho} DGR_{t-2}, \quad \text{etc}$$

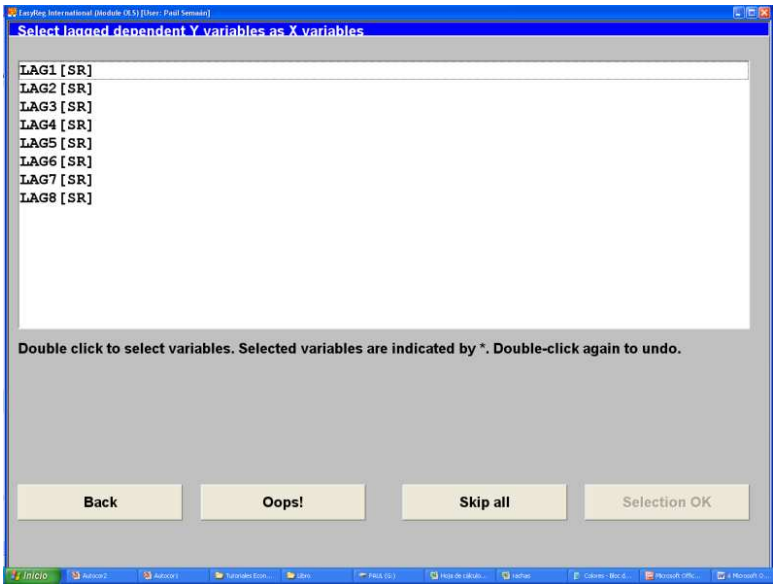
3. Estime el modelo:

$$SR^*_t = \beta^*_0 + \beta_1 DGR^*_{t-1} + \beta_2 r^*_t + \beta_3 BC^*_t + \upsilon_t \tag{6.9}$$

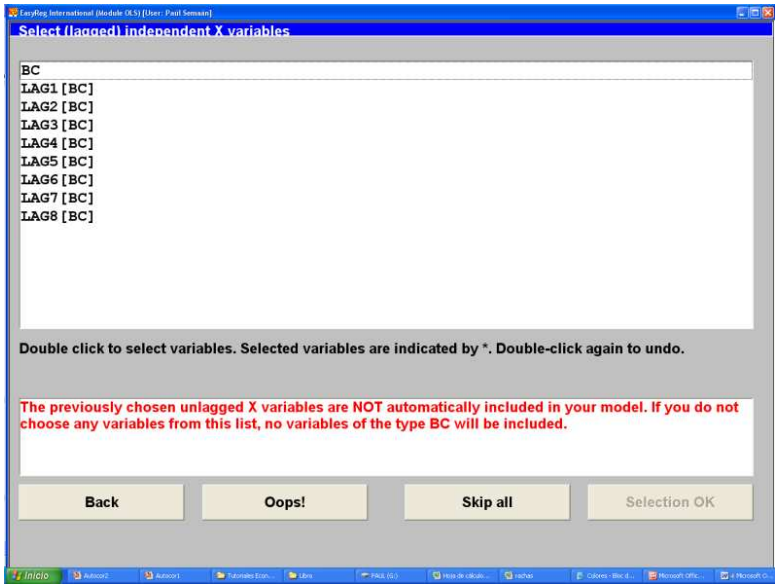
donde $\beta^*_0 = \beta_0 (1 - \hat{\rho})$

Para correr la regresión del paso 1 haga clic en "Menu / Single equation Models/Linear regression models"

A continuación, haga doble clic en las variables “SR” “BC” “DGR”, “r”,. Haga clic en el botón “Selection OK”, después en “No” y posteriormente en “Continue”. Escoja la variable dependiente “SR” y posteriormente haga clic en “Continue” dos veces. En la ventana de selección de variables independientes, haga clic en “Selection OK”, verá la siguiente ventana



Haga nuevamente clic en “LAG1[SR]” y posteriormente en el botón “Selection OK”. Observará la siguiente ventana



Ahora, seleccione “BC”, luego “LAG1[BC]” y posteriormente haga clic en el botón "Selection OK". Repita este procedimiento con las otras variables independientes. Encontrará que $\rho = 0{,}681001$

Ya podemos efectuar el segundo paso, es decir, necesitamos crear las siguiente variables $SR_t^* = SR_t - \hat{\rho}SR_{t-1}$, $DGR_{t-1}^* = DGR_{t-1} - \hat{\rho}DGR_{t-2}$, etc. Para esto vaya al menú “Transform variables” y haga clic en el botón “Time series transformations”. Haga clic en el botón “Lag: x(t-m)” y nuevamente en el botón “Lag: x(t-m)”. Haga doble clic en las variables “SR”, “BC”, “DGR”, “r” y haga clic en el botón “Selection OK” y en el botón “OK”. Las variables han sido creadas bajo el nombre “LAG1[SR]”, “LAG1[BC]”, etc. Repita este procedimiento para la variable “DGR” ya que necesitamos que esta se encuentre rezagada dos periodos Cree las variables $SR_t^* = SR_t - \hat{\rho}SR_{t-1}$, $DGR_{t-1}^* = DGR_{t-1} - \hat{\rho}DGR_{t-2}$, etc. (Emplee la opción “linear combination of variables”).

Ahora, estime el modelo $SR_t^* = \beta_0^* + \beta_1 DGR_{t-1}^* + \beta_2 r_t^* + \beta_3 BC_t^* + \upsilon_t$, obtendrá el siguiente resultado (Note que este modelo ya no tiene ningún problema de autocorrelación).[12]

[12]Es importante realizar todas las pruebas de nuevo para este modelo. Sólo de esta manera estaremos seguros de que el problema fue solucionado. En este caso el lector podrá constatar que todas las pruebas muestran que el modelo ya está libre de problemas

Recuadro 6.7 Output de EasyReg Modelo Corregido

Dependent variable:
Y = SR*
Characteristics:
SR*
First available observation = 2(=1965)
Last available observation = 41(=2004)
First chosen observation = 3(=1966)
Last chosen observation = 41(=2004)
Number of usable chosen observations: 39
Subsample characteristics:
Minimum value: -4.3642633E+000
Maximum value: 2.9600183E+000
Sample mean: -7.2985003E-001

X variables:
X(1) = BC*
X(2) = DGR*
X(3) = r*
X(4) = 1

Model:
Y = b(1)X(1) +.....+ b(4)X(4) + U,
where U is the error term, satisfying
E[U|X(1),...,X(4)] = 0.

OLS estimation results

Parameters	Estimate	t-value (S.E.) [p-value]	H.C. t-value (H.C. S.E.) [H.C. p-value]
b(1)	0.11661	3.006 (0.03879) [0.00265]	3.032 (0.03845) [0.00243]
b(2)	0.02660	0.547 (0.04866) [0.58454]	0.804 (0.03308) [0.42121]
b(3)	-0.00274	-0.068 (0.04036) [0.94592]	-0.055 (0.04937) [0.95577]
b(4)	1.04803	1.884 (0.55618) [0.05952]	1.778 (0.58955) [0.07546]

Notes:
1: S.E. = Standard error
2: H.C. = Heteroskedasticity Consistent. These t-values and
standard errors are based on White's heteroskedasticity
consistent variance matrix.
3: The two-sided p-values are based on the normal approximation.
Effective sample size (n): 39
Variance of the residuals: 1.467608
Standard error of the residuals (SER): 1.211449
Residual sum of squares (RSS): 51.366291
(Also called SSR = Sum of Squared Residuals)

Total sum of squares (TSS): 68.578446
R-square: 0.2510

Adjusted R-square: 0.1868

Overall F test: F(3,35) = 3.91
p-value = 0.01652

Significance levels:	10 %	5 %
Critical values:	2.25	2.87
Conclusions:	reject	reject

Test for first-order autocorrelation:
Durbin-Watson test = 2.047852
REMARK: A better way of testing for autocorrelation
is to specify AR errors and then test the null
hypothesis that the AR parameters are zero.

Jarque-Bera/Salmon-Kiefer test = 0.676389
Null hypothesis: The errors are normally distributed
Null distribution: Chi-square(2))
p-value = 0.71306

Significance levels:	10 %	5 %
Critical values:	4.61	5.99
Conclusions:	accept	accept

Breusch-Pagan test = 5.152737
Null hypothesis: The errors are homoskedastic
Null distribution: Chi-square(3)
p-value = 0.16095

Significance levels:	10 %	5 %
Critical values:	6.25	7.81
Conclusions:	accept	accept

Information criteria:
Akaike: 4.80549E-01
Hannan-Quinn: 5.41766E-01
Schwarz: 6.51170E-01

Estos resultados se resumen en el cuadro 6.5

Después de efectuar esta corrección por el método de diferencias generalizadas, es importante chequear si el modelo con la corrección está libre de problemas. En este caso encontramos los siguientes resultados para las 4 pruebas discutidas.

Para la prueba de Rachas se obtiene un número de rachas (k) igual a 22, cantidad que se halla entre el intervalo al 95 % de confianza ([13 , 26]). Por lo tanto, no existe suficiente evidencia para rechazar la hipótesis nula de errores totalmente aleatorios con un nivel de confianza del 95 %. De igual manera, al realizar la prueba de Durbin Watson para el modelo corregido, obtenemos un d_l = 1,328 y un d_u = 1,658 y con un valor DW = 2,048 no es posible rechazar la hipótesis nula de no autocorrelación.

Al realizar la prueba de Breusch-Godfrey no es posible rechazar la hípótesis nula de no autocorrelación en ninguno de los procesos examinados (cuadro 6.3). Posteriormente, en las pruebas de Box-Pierce y Ljung-Box tampoco se puede rechazar la hipótesis nula de no autocorrelación en los diferentes rezagos (cuadro 6.4).

Por lo tanto, ninguna de las pruebas realizadas muestra evidencia de un proceso autorregresivo de los errores.

Cuadro 6.3: Resultados de las pruebas de Breusch-Godfrey

	Proceso Autorregresivo del Error				
	$AR(1)$	$AR(2)$	$AR(3)$	$AR(4)$	$AR(5)$
R^2	0.0091	0.0659	0.0854	0.1216	0.1302
LM	0.3367	2.3065	2.8182	3.7696	3.7758

Cuadro 6.4: Resultados de las pruebas de Box-Pierce y Ljung-Box

	Estadístico	
Rezago	Box-Pierce	Ljung-Box
1	0.03	0.03
2	2.33	2.33
3	2.58	2.58
4	3.21	3.21
5	3.23	3.23
6	7.52	7.52
7	7.95	7.95
8	12.44	12.44
9	13.69	13.69
10	13.74	13.74

Cuadro 6.5: Estimación del modelo 6.9

	Variable dependiente SR_t Estadísticos t entre paréntesis
	Ecuación 6.9 D.G
Constante	3.2853
	N.A.
DGR_{t-1}	0.0.266
	(0.547)
BC_t	0.1166
	(3.006)***
r_t	-0.00274
	(-0.068)
R^2	0.2510
R2 Ajustado	0.1868
DW	2.0478
F-Global	3.91**
No. de obs.	39

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

D.G: Diferencias generalizadas

Es importante aclarar que para calcular el intercepto se puede despejar de la siguiente manera:

$$\hat{\beta}_0 = \frac{\hat{\beta}_0^*}{(1-\hat{\rho})} = \frac{1{,}048}{(1-0{,}681)} = 3{,}285$$

Además, el estadístico t asociado al intercepto no aplica para el modelo, ya que el intercepto encontrado mediante el modelo corregido es difernte al del modelo original.

Por lo tanto, podemos darnos cuenta que $\hat{\delta}$ no es significativo, lo que permite concluir que los incrementos que presenta la deuda pública como porcentaje del PIB (DGR_t) no afectan el superávit primario como porcentaje del PIB (SR_t). De esta manera, se cumple la condición de sostenibilidad fiscal para el periodo de estudio en el caso colombiano.

6.5. Ejercicios

El gobierno de un pequeño país caribeño está interesado en conocer la relación existente entre el ingreso disponible y el consumo privado, con el fin de establecer la política impositiva que regirá en el año siguiente. Se cuenta con datos sobre el consumo privado y la renta disponible recolectada desde el primer trimestre de 1959 hasta el tercer trimestre del año 2002, ambos valores medidos en millones de moneda local. Los datos se encuentran en el archivo auto.xls.

De acuerdo con la siguiente información responda:

1. Estime el modelo que explica la situación. Reporte sus resultados en una tabla.

2. Efectúe el análisis gráfico de los errores estimados. ¿Qué tipo de problema puede intuir a partir de este análisis? Explique.

3. Realice las pruebas que considere necesarias para determinar la existencia o no de un problema de autocorrelación en el modelo. Especifique siempre las hipótesis que sustentan cada prueba y muestre claramente la conclusión a la que llega.

4. Según las conclusiones que extrajo del punto anterior:

   a) ¿A qué conclusión llega? ¿explique por medio de la teoría econométrica cómo solucionaría el problema si es que este existe?

   b) Ahora, demuestre que el problema ha desaparecido (realice las pruebas pertinentes). Además estime el nuevo modelo e intérprete los coeficientes estimados (siempre y cuando haya encontrado un problema econométrico en el modelo original)

6.6. Apéndice

Apéndice 6.1 Estadísticos de la prueba de Rachas

Cuadro 6.6: Límite inferior de la prueba de Rachas. Nivel de confianza del 95 %

	Número de errores negativos (N_-)																			
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Número de errores negativos (N_+)	2											2	2	2	2	2	2	2	2	2
	3					2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
	4				2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
	5			2	2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
	6		2	2	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
	7		2	2	3	3	3	4	4	5	5	5	5	5	6	6	6	6	6	6
	8		2	3	3	3	4	4	5	5	5	6	6	6	6	6	7	7	7	7
	9		2	3	3	4	4	5	5	5	6	6	6	7	7	7	7	8	8	8
	10		2	3	3	4	5	5	5	6	6	7	7	7	7	8	8	8	8	9
	11		2	3	4	4	5	5	6	6	7	7	7	8	8	8	9	9	9	9
	12	2	2	3	4	4	5	6	6	7	7	7	8	8	8	9	9	9	10	10
	13	2	2	3	4	5	5	6	6	7	7	8	8	9	9	9	10	10	10	10
	14	2	2	3	4	5	5	6	7	7	8	8	9	9	9	10	10	10	11	11
	15	2	3	3	4	5	6	6	7	7	8	8	9	9	10	10	11	11	11	12
	16	2	3	4	4	5	6	6	7	8	8	9	9	10	10	11	11	11	12	12
	17	2	3	4	4	5	6	7	7	8	9	9	10	10	11	11	11	12	12	13
	18	2	3	4	5	5	6	7	8	8	9	9	10	10	11	11	12	12	13	13
	19	2	3	4	5	6	6	7	8	8	9	10	10	11	11	12	12	13	13	13
	20	2	3	4	5	6	6	7	8	9	9	10	10	11	12	12	13	13	13	14

Funte: Swed y Eisenhart (1943)

Cuadro 6.7: Límite superior de la prueba de Rachas. Nivel de confianza del 95 %

	Número de errores negativos (N_-)																			
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Número de errores negativos (N_+)	2																			
	3																			
	4				9	9														
	5			9	10	10	11	11												
	6			9	10	11	12	12	13	13	13	13								
	7				11	12	13	13	14	14	14	14	15	15	15					
	8				11	12	13	14	14	15	15	16	16	16	16	17	17	17	17	17
	9					13	14	14	15	16	16	16	17	17	18	18	18	18	18	18
	10					13	14	15	16	16	17	17	18	18	18	19	19	19	20	20
	11					13	14	15	16	17	17	18	19	19	19	20	20	20	21	21
	12					13	14	16	16	17	18	19	19	20	20	21	21	21	22	22
	13						15	16	17	18	19	19	20	20	21	21	22	22	23	23
	14						15	16	17	18	19	20	20	21	22	22	23	23	23	24
	15						15	16	18	18	19	20	21	22	22	23	23	24	24	25
	16							17	18	19	20	21	21	22	23	23	24	25	25	25
	17							17	18	19	20	21	22	23	23	24	25	25	26	26
	18							17	18	19	20	21	22	23	24	25	25	26	26	27
	19							17	18	20	21	22	23	23	24	25	26	26	27	27
	20							17	18	20	21	22	23	24	25	25	26	27	27	28

Funte: Swed y Eisenhart (1943)

Apéndice 6.2 Estadísticos de la prueba de Durbin-Watson

Cuadro 6.8: Límites inferior (d_l) y superior (d_u) para la prueba Durbin-Watson con un nivel de confianza del 95 %

	Número de coeficientes estimados[1]									
Número de	2		3		4		5		6	
observaciones	dl	du	dl	du	dl	du	dl	du	dl	du
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.45	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Funte: Durbin y Watson (1951)

[1] Incluyendo la constante

Cuadro 6.9: Límites inferior (d_l) y superior (d_u) para la prueba Durbin-Watson con un nivel de confianza del 99 %

	Número de coeficientes estimados[1]									
Número de	2		3		4		5		6	
observaciones	dl	du	dl	du	dl	du	dl	du	dl	du
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
16	0.84	1.09	0.74	1.25	0.63	1.44	0.53	1.66	0.44	1.90
17	0.87	1.10	0.77	1.25	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.80	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.83	1.26	0.74	1.41	0.65	1.58	0.56	1.77
20	0.95	1.15	0.86	1.27	0.77	1.41	0.68	1.57	0.60	1.74
21	0.97	1.16	0.89	1.27	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.91	1.28	0.83	1.40	0.75	1.54	0.66	1.69
23	1.02	1.19	0.94	1.29	0.86	1.40	0.77	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.80	1.53	0.72	1.66
25	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.85	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.51	0.81	1.63
28	1.10	1.24	1.04	1.32	0.97	1.41	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.85	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.08	1.34	1.02	1.42	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.05	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.43	1.01	1.51	0.95	1.59
35	1.19	1.31	1.14	1.37	1.08	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.16	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.12	1.45	1.07	1.52	1.02	1.58
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
45	1.29	1.38	1.24	1.42	1.20	1.48	1.16	1.53	1.11	1.58
50	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.31	1.57	1.28	1.61
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.39	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.43	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.56	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.42	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65

Fuente: Durbin y Watson (1951)

[1] Incluyendo la constante

Apéndice 6.3 Demostración de la insesgadez de los estimadores en presencia de autocorrelación

En presencia de autocorrelación los estimadores MCO siguen siendo insesgados. Esta afirmación se puede demostrar fácilmente. Sin perder generalidad consideremos un modelo lineal con un término de error homoscedástico y con autocorrelación de orden uno. Es decir,

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

Donde $E[\varepsilon_t] = 0$, $Var[\varepsilon_t] = \sigma_\varepsilon^2$ y $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$, $\forall t$. Ahora determinemos si $\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$ sigue siendo insesgado o no. Así,

$$E\left[\hat{\beta}\right] = E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}E[\mathbf{y}]$$

$$E\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}E[\mathbf{X}\beta + \varepsilon] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta + \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}E[\varepsilon]$$

$$E\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta = I \bullet \beta$$

$$E\left[\hat{\beta}\right] = \beta$$

Apéndice 6.4 Sesgo de la matriz de varianzas y covarianzas en presencia de autocorrelación

En presencia de autocorrelación el estimador de la matriz de varianzas y covarianzas de MCO ($\widehat{Var\left[\hat{\beta}\right]} = s^2\left(\mathbf{X^T X}\right)$) es sesgado. Es más el estimador MCO para los coefcientes ($\hat{\beta} = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}$) no es eficiente; es decir no tiene la mínima varianza posible. Esta afirmación se puede demostrar facilmente.

Continuando con el modelo considerado en el Apéndice anterior (error con una autorrelación de orden uno), en este caso tenemos que:

$$Var[\varepsilon] = E\left[\varepsilon^T\varepsilon\right] = \Omega = \sigma_\varepsilon^2\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho & 1 \end{bmatrix} \tag{6.10}$$

Ahora podemos calcular la varianza de los estimadores MCO. Es decir,

$$Var\left[\hat{\beta}\right] = Var\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}\right] \tag{6.11}$$

.

Por tanto tendremos que

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}Var[\mathbf{y}]\,\mathbf{X}\left(\mathbf{X^T X}\right)^{-1} \tag{6.12}$$

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}Var[\varepsilon]\,\mathbf{X}\left(\mathbf{X^T X}\right)^{-1} \tag{6.13}$$

$$Var\left[\hat{\beta}\right] = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T} \Omega \mathbf{X} \left(\mathbf{X^T X}\right)^{-1} \tag{6.14}$$

Por tanto la varianza no es la mínima posible. Y por otro lado, al emplear el estimador MCO para la matriz de varianzas y covarianzas de los betas ($\widehat{Var\left[\hat{\beta}\right]} = s^2 \left(\mathbf{X^T X}\right)^{-1}$) en presencia de autocorrelación se obtendrá un estimador cuyo valor esperado no es igual a la varianza real; es decir, será insesgado.

Parte III

Otros modelos econométricos

CAPÍTULO 7

Modelos Logit y Probit

Objetivos del capítulo

Al finalizar este capítulo, el estudiante estará en capacidad de:

- Estimar mediante el uso de EasyReg un modelo Logit.
- Estimar mediante el uso de EasyReg un modelo Probit.
- Calcular el efecto marginal promedio de cada variable de un modelo estimado (Logit o Probit) a partir de los resultados calculados por EasyReg.

7.1. Introducción

Hasta el momento hemos discutido modelos cuya variable dependiente es una variable continua (o que se supone continua),[1] en este capítulo centraremos la atención en modelos cuya variable dependiente es una variable dummy (dicotómica). Estos modelos permiten estudiar variables dependientes como por ejemplo: i) si un individuo pertenecerá o no a la PEA, ii) si un individuo será cliente o no, iii) si un individuo pagará o no.

Una posibilidad para estimar este tipo de modelos es emplear el modelo lineal y estimarlo por MCO. Esta aproximación de emplear el modelo lineal con una variable dependiente se conoce como el Modelo de Probabilidad Lineal (MPL). Lastimosamente esta aproximación de estimar por MCO el MPL tiene varios problemas, unos solucionables y otros no.

Los problemas más comunes de estimar modelos con variables dummy como dependientes por MCO son:

1. El error del modelo presenta problemas de heteroscedasticidad.[2] De hecho se sabe que en este caso: $\sigma_i^2 = E[\mathbf{y_i}|\mathbf{x}_i](1 - E[\mathbf{y_i}|\mathbf{x}_i])$, donde $\mathbf{x}_i$ representa un vector fila de dimensiones $(1 \times k)$ que recoge las características del individuo i.[3].

2. El modelo estimado no garantiza que los valores estimados para la variable dependiente estén en el intervalo [0,1]. Esto implica, por ejemplo, varianzas negativas.

El primero de los problemas se puede solucionar empleando el método de Mínimos Cuadrados Ponderados (MCP) discutidos en el capítulo 5.[4] El segundo problema no tiene una solución fácil y hace que este modelo no sea muy usado en la actualidad. Para entender en detalle este problema empleemos un ejemplo.

Supongamos que se desea explicar el hecho de que un individuo pertenezca o no a la población económicamente activa (PEA) por medio de sólo una variable: los años de educación ($X_{2,i}$), Es decir, se tiene que $\mathbf{x}_i = [1, X_{2i}]$, por tanto:

$$\begin{aligned} y_i &= \beta^T \mathbf{x}_i^T + \varepsilon_i \\ y_i &= \beta_1 + \beta_2 X_{2,i} + \varepsilon_i \end{aligned} \tag{7.1}$$

Así, para estimar el modelo se tendrá que recoger información de individuos que pertenecen a la PEA ($y_i = 1$) y aquellos que no ($y_i = 0$) y los respectivos años de educación de cada individuo. Al graficar estos datos tendremos algo parecido al panel a) de la figura 7.1. Noten que sólo se tendrán observaciones de la variable dependiente con valores de 1 y 0. El método de MCO lo que intentará hacer es encontrar una línea recta que minimice la suma de la distancia cuadrada entre los puntos

[1] Como por ejemplo las unidades demandadas de carros; que si bien no es una variable continua, se supone como tal.

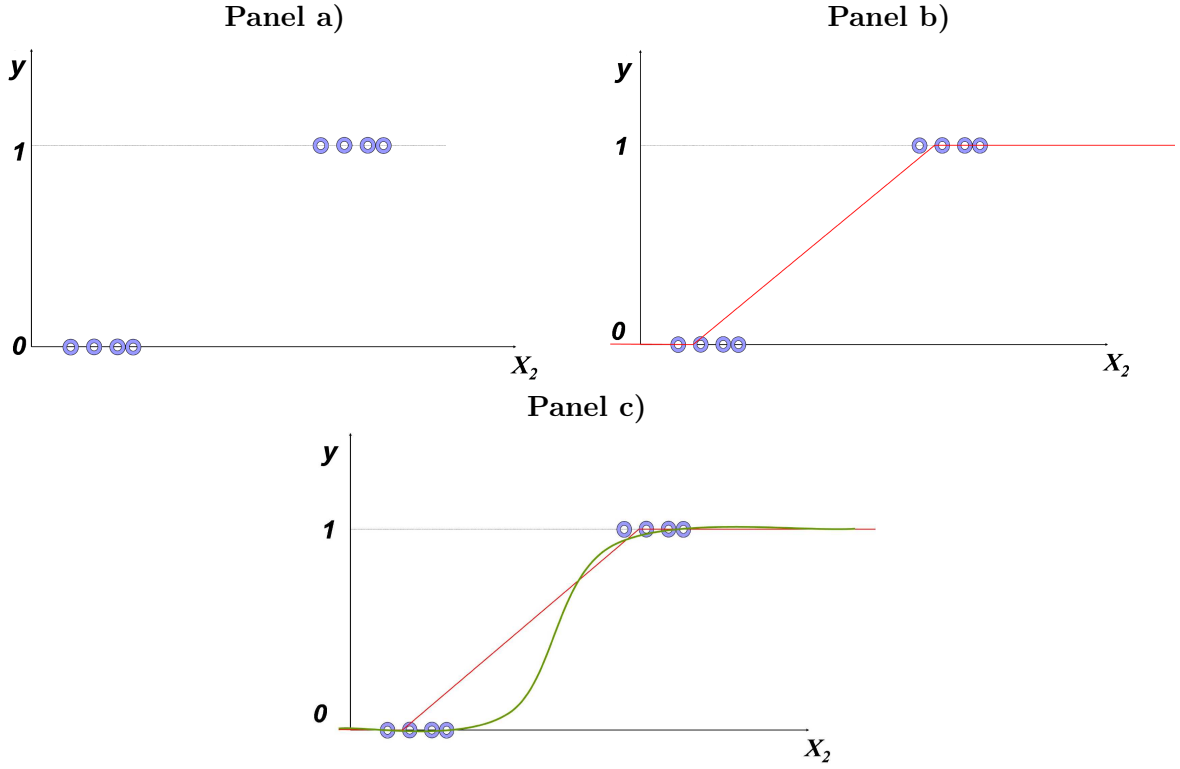
[2] En el apéndice 7.1 se presenta una prueba de esta afirmación.

[3] Es decir, corresponde a la fila i de la matriz $\mathbf{x}_i$.

[4] Es importante anotar que para emplear el método de MCP será necesario emplear como variable ponderadora $\omega_i^2 = E[\mathbf{y_i}|\mathbf{x}_i](1 - E[\mathbf{y_i}|\mathbf{x}_i])$. Naturalmente, en el caso del MPL se tiene que $E[y_i|\mathbf{x}_i] = \beta^T \mathbf{x}_i^T$. Dado que el vector β no es conocido es imposible calcular ω_i . Esto hace necesario estimar esta variable ponderadora en un primer paso. Así, emplear el método de MCP implicará un primer paso para estimar el modelo original por MCO para obtener $\hat{\beta}_{OLS}$. Con estos coeficientes estimados se puede encontrar la variable ponderadora $W_i = \sqrt{\hat{y}_i(1 - \hat{y}_i)}$, donde $\hat{y}_i = \left(\hat{\beta}_{OLS}\right)^T \mathbf{x}_i^T$. El segundo paso será transformar los datos dividiéndolos por W_i. El tercer paso implica estimar el modelo por MCO con las variables transformadas.

observados y la línea estimada. Así, se obtendrá una línea como la presentada en el panel b) de ese mismo gráfico. Noten que el modelo estimado ($\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2,i}$) no acota los valores estimados para la variable dependiente al intervalo [0,1] .

Figura 7.1: Modelo lineal con variable dummy como dependiente



Es más, este gráfico permite entender que un modelo lineal no podrá explicar de manera adecuada este tipo de modelos en los que la variable dependiente solo toma dos valores. De hecho parece más adecuado emplear modelos no lineales para ajustar los datos, tal como se muestra en el panel c) de la figura 7.1. Dos tipos de modelos no lineales que se pueden emplear en estos casos son los modelos Probit y Logit.

Una manera de asegurar que los valores estimados del modelo se mantengan en el intervalo [0,1] es emplear una forma funcional acorde. Esto se garantiza empleando funciones de probabilidad acumulativa. Para entender cómo emplear estas funciones de probabilidad, empleemos el mismo ejemplo anterior, pero permitamos la posibilidad de más variables explicativas.

Es decir, suponga que se quiere explicar la variable y_i tal como se definió en nuestro ejemplo, empleando $k-1$ variables explicatorias y una constante. Así, emplearemos el vector fila $\mathbf{x}_i$ de dimensiones $(1 \times k)$, que recoge las características del individuo i para explicar la probabilidad de pertenecer a la PEA.

Un individuo que está decidiendo si participa en la PEA decidirá si participa en la PEA o no dependiendo de si el beneficio neto de participar en ella es positivo o no. Así, si el beneficio neto de participar en la PEA es positivo, el individuo participará y si éste es negativo, entonces el individuo no participará. Por ahora definamos el beneficio neto del individuo i como Z_i. Así, tendremos que:

$$y_i = \begin{cases} 1 & si \quad Z_i > 0 \\ 0 & si \quad Z_i \leq 0 \end{cases} \tag{7.2}$$

Pero, ¿de qué depende ese beneficio neto? Supongamos que éste es función de las características del individuo recolectadas en el vector $\mathbf{x_i}$. Es decir:

$$\mathbf{Z_i} = \beta^T \mathbf{x}_i^T + \varepsilon_i$$

Donde ε_i es un término aleatorio de error. Así, para unas características dadas para el individuo i tendremos que el valor esperado del beneficio neto será:

$$E\left[Z_i \left| \mathbf{x}_i \right.\right] = \bar{Z}_i = \beta^T \mathbf{x_i^T}$$

Recapitulando, las características del individuo i ($\mathbf{x}_i$) afectan el beneficio neto (Z_i) y éste a su vez explica si el individuo pertenece a la PEA o no; es decir, y_i . Ahora bien, noten que la distribución del beneficio neto, dadas unas características del individuo i ($\mathbf{x}_i$), dependerá de la distribución del término de error ε_i.

Por conveniencia, podemos emplear el hecho que $Z_i = \beta^T \mathbf{x}_i^T + \varepsilon_i > 0$ implica que $\varepsilon_i > -\beta^T \mathbf{x}_i^T = -Z_i$ para reescribir la ecuación 7.3 de la siguiente manera:

$$y_i = \begin{cases} 1 & si \quad -Z_i < 0 \\ 0 & si \quad -Z_i \geq 0 \end{cases} \tag{7.3}$$

En el panel a) de la figura 7.2 se puede observar un esquema de un individuo con una relativa baja probabilidad de pertenecer a la PEA, dadas sus características; mientras que en el panel b) se puede observar un individuo con una probabilidad relativamente alta. Es importante anotar que la probabilidad está expresada por el área bajo la curva de la función de distribución y dependerá de la forma de ésta. Finalmente, cabe anotar que la función que determina el área bajo la curva a la izquierda de cualquier valor determinado de una función de distribución se conoce como la función de distribución acumulativa. Es decir,

$$F\left(Z_i\right) = P\left(Z_i \leq c\right) = \int_{-\infty}^{c} f\left(t\right) dt$$

En otras palabras se tendrá que:

$$P\left[Y_i = 0 \left| \mathbf{x}_i \right.\right] = F\left(-Z_i\right) = F\left(-\beta^T \mathbf{x_i^T}\right)$$
$$P\left[Y_i = 1 \left| \mathbf{x}_i \right.\right] = 1 - F\left(-Z_i\right) = 1 - F\left(-\beta^T \mathbf{x_i^T}\right)$$

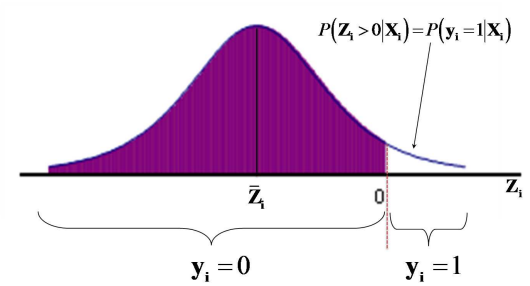
Precisamente esta función de probabilidad acumulativa para distribuciones simétricas como la normal presenta una forma de "S" como la mostrada en el panel c) de la figura 7.1

Por otro lado, lastimosamente Z_i es una variable que los investigadores no pueden observar directamente, si bien se puede realizar una inferencia sobre esta al observar una variable relacionada, como lo es en este caso la variable dummy y_i. A este tipo de variables se les denomina *variables latentes*. Es decir, una variable latente es una variable no observable directamente, pero para la cual se puede inferir su valor a partir de una variable relacionada.

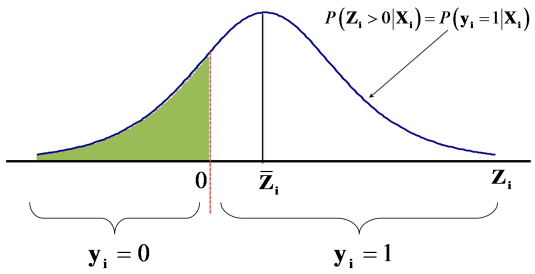
Retornando a nuestro problema, lo que se desea es encontrar los coeficientes β que en últimas refleja el impacto de cada una de las variables explicativas sobre la probabilidad de que y_i sea uno. Es

Figura 7.2: Distribución del beneficio neto (de la variable latente)

Panel a) Individuo con baja probabilidad de pertenecer a la PEA



Panel b) Individuo con alta probabilidad de pertenecer a la PEA



decir, dada una muestra de tamaño n y sus correspondientes características ($\mathbf{x}_i$) se desean encontrar los $\beta's$ tal que $P\left[y_i = 1\left|\mathbf{x}_i\right.\right] = 1 - F\left(-\mathbf{Z_i}\right) = 1 - F\left(-\beta^T\mathbf{x}_i^T\right)$ si el individuo pertenece a la PEA y $P\left[y_i = 0\left|\mathbf{x}_i\right.\right] = F\left(-Z_i\right) = F\left(-\beta^T\mathbf{x}_i^T\right)$ si el individuo no pertenece a la PEA.

A la probabilidad de que la i-esima observación fuera igual a lo realmente observado se le denomina verosimilitud (en inglés likelihood). Así, la verosimilitud para un individuo que pertenece a la PEA es $P\left[y_i = 1\left|\mathbf{x}_i\right.\right]$ y para los que no pertenecen a la PEA es $P\left[y_i = 0\left|\{\mathbf{x}_i\right.\right]$. La función que recoge la verosimilitud (probabilidad) de que ocurra toda la muestra se conoce como la *función de verosimilitud*, es decir:

$$\begin{aligned} L(\beta) = & P\left(y\right) = \prod_{\forall i\, tal\, que\, y_i=1} P\left(y_i = 1\right) \prod_{\forall i\, tal\, que\, y_i=0} P\left(y_i = 0\right) \\ = & \prod_{\forall i\, tal\, que\, y_i=1} 1 - F\left(-Z_i\right) \prod_{\forall i\, tal\, que\, y_i=0} F\left(-Z_i\right) \end{aligned}$$

Ahora bien, para encontrar los coeficientes β de este tipo de problemas ya no tiene mucho sentido emplear el método de mínimos cuadrados estudiado hasta el momento. De hecho tiene más sentido tratar de encontrar el vector de coeficientes (β) tal que la función de verosimilitud sea maximizada. A este método se le denomina **Método de Máxima Verosimilitud**; método que implica encontrar los coeficientes tal que el chance de observar la muestra con que se cuenta sea el máximo posible, bajo el supuesto que la información sigue una distribución dada por $F\left(\bullet\right)$.

En otras palabras, los estimadores de máxima verosimilitud corresponden a los coeficientes $\hat{\beta}^{MV}$ tal que:

$$\begin{aligned} \underset{\beta}{Max}\, L\left(\beta\right) = \underset{\beta}{Max}\left[P\left(y\right)\right] = & \underset{\beta}{Max}\left[\prod_{\forall i\, tal\, que\, y_i=1} P\left(y_i = 1\right) \prod_{\forall i\, tal\, que\, y_i=0} P\left(y_i = 0\right)\right] \\ = & \underset{\beta}{Max}\left[\prod_{\forall i\, tal\, que\, y_i=1} 1 - F\left(-Z_i\right) \prod_{\forall i\, tal\, que\, y_i=0} F\left(-Z_i\right)\right] \end{aligned} \tag{7.4}$$

En general, es mucho más fácil maximizar el logaritmo de la función de máxima verosimilitud (conocida en inglés con el nombre de log likelihood function) que la expresión 7.4. Es decir, en la práctica el problema que se resuelve es:

$$\underset{\beta}{Max} \ln\left(L\left(\beta\right)\right) = \underset{\beta}{Max}\left[\ln\left(P\left(Y\right)\right)\right] \tag{7.5}$$

Dado que la función logarítmica es una transformación monotónica, el vector β que resuelva el problema 7.4 también solucionará el problema 7.5. Es importante resaltar que es difícil encontrar, y en algunos casos no existe, una fórmula que siempre que se aplique dé solución al problema de Máxima Verosimilitud,[5] por tanto el proceso de maximización se resuelve de manera numérica empleando computadores.

Los estimadores de máxima verosimilitud son consistentes y siguen una distribución aproximadamente igual a la normal en muestras grandes. Así mismo a medida que la muestra crece estos estimadores serán eficientes. Este resultado hace que la inferencia sobre los coeficientes, tanto individual como conjunta, sea igual que lo estudiado para los estimadores MCO.

7.2. Modelo Probit

Como se discutió anteriormente, cuando se cuenta con variables dicotómicas como variables dependientes, la probabilidad de observar un valor de uno o cero para una observación depende de la función de distribución que se asuma y más específicamente de su respectiva función de probabilidad acumulativa.

Una función acumulativa que tiene una forma de "S" como la sugerida en el panel c) de la figura 7.1 es la correspondiente a la distribución normal (ver figura 7.3). En este caso si suponemos que la función de distribución acumulativa corresponde a una distribución estándar normal tendremos un modelo de elección binario denominado *modelo Probit*, lo cual implica que:

$$F(Z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-Z_i} e^{-\frac{t^2}{2}} dt = \Phi(-Z_i) \tag{7.6}$$

Entonces los estimadores de Máxima Verosimilitud implicarán resolver el siguiente problema

$$\begin{aligned} \underset{\beta}{Max} \left(\ln \left[\prod_{\forall i \, tal \, que \, y_i=1} \left(1 - \tfrac{1}{\sqrt{2\pi}} \int_{-}^{-Z_i} e^{-\frac{t^2}{2}} dt \right) \prod_{\forall i \, tal \, que \, y_i=0} \left(\tfrac{1}{\sqrt{2\pi}} \int_{-}^{-Z_i} e^{-\frac{t^2}{2}} dt \right) \right] \right) \\ = \underset{\beta}{Max} \left(\ln \left[\prod_{\forall i \, tal \, que \, y_i=1} (1 - \Phi(-Z_i)) \prod_{\forall i \, tal \, que \, y_i=0} (\Phi(-Z_i)) \right] \right) \end{aligned} \tag{7.7}$$

Como se mencionó anteriormente, los computadores pueden encontrar fácilmente por medio de métodos numéricos el vector de coeficientes β^{MV} que resuelve este problema y su correspondiente matriz de varianzas y covarianzas.

Un aspecto importante para resaltar en la estimación de un modelo Probit es que los coeficientes β no tienen interpretación por si solos. Recuerde que una práctica común para interpretar los coeficientes de un modelo es encontrar la derivada del valor esperado de la variable dependiente con respecto a las variables explicativas. En este caso tenemos que:

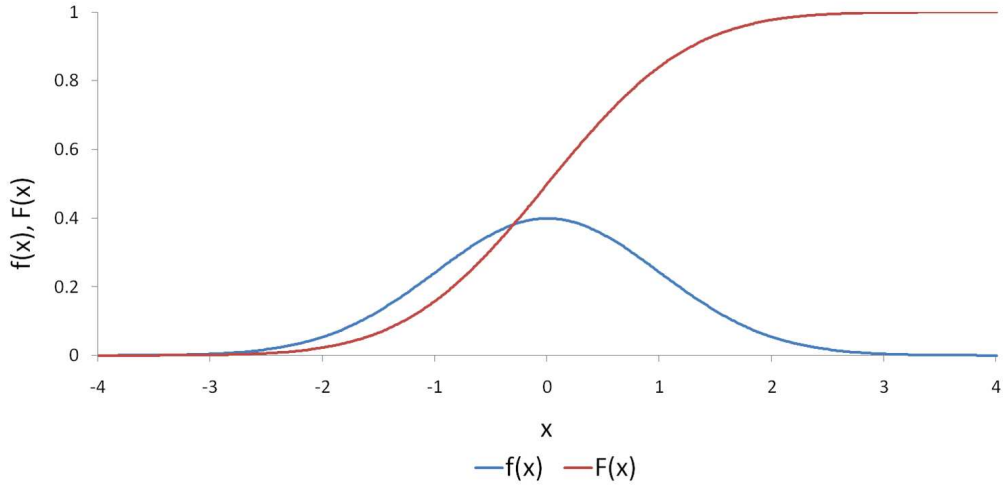
$$E[y_i | \mathbf{x}_i] = 1 \bullet (1 - \Phi(-\mathbf{Z_i})) + 0 \bullet \Phi(-\mathbf{Z_i}) = 1 - \Phi(-\mathbf{Z_i})$$

Y por tanto

$$\frac{\partial E[y_\mathbf{i} | \mathbf{x_{i}}]}{\partial \mathbf{X_j}} = -\frac{\partial \Phi(-\mathbf{Z_i})}{\partial \mathbf{X_j}} \beta_j = \frac{1}{\sqrt{2\pi}} e^{-\frac{\mathbf{z_i^2}}{2}} \beta_j = \phi(\mathbf{Z_i}) \beta_j$$

[5] Recordemos que en el caso de los estimadores MCO, sí existe una fórmula $(X^T X)^{-1} X^T Y$ que garantiza que al aplicarla la suma de los residuos al cuadrado será mínima. Esa fórmula se denomina estimador.

Figura 7.3: Función de distribución estándar normal ($f(x)$) y su correspondiente distribución acumulativa ($F(x)$)



Esto implica que el efecto de un aumento en una unidad en la variable X_j provocará un cambio de $\phi(Z_i)\beta_j$ en la probabilidad de que $y_i = 1$. En otras palabras, el efecto marginal de X_j no es el mismo para todos los individuos.

Si la muestra es relativamente grande, puede ser muy complicado analizar el efecto marginal de cada una de las variables explicativas para cada uno de los elementos de la muestra. En ese caso una práctica común es calcular todos los efectos marginales para la muestra y calcular el promedio de este. Es decir:

$$\frac{\sum_{i=1}^{n} \phi(Z_i)\beta_j}{n}$$

7.3. Modelo Logit

Otra opción para estimar un modelo con variable dicotómica como variable dependiente es suponer que la función acumulativa de distribución proviene de una distribución logística. En este caso la forma de la distribución logística es relativamente parecida a la distribución normal (la del modelo Probit).

En este caso tendremos que:

$$F(Z_i) = \frac{e^{-Z_i}}{1+e^{-Z_i}} = \Lambda(-Z_i)$$

Entonces los estimadores de Máxima Verosimilitud implicarán resolver el siguiente problema

$$\underset{\beta}{Max}\left(\ln\left[\prod_{\forall i\, tal\, que\, y_i=1}(1-\Lambda(-Z_i))\prod_{\forall i\, tal\, que\, y_i=0}(\Lambda(-Z_i))\right]\right)$$

Para el modelo Logit, al igual que en el Probit, los coeficientes β no tienen interpretación por sí solos. En este caso tenemos que:

$$E[y_i|\mathbf{x}_i] = 1 \bullet (1-\Lambda(-\mathbf{Z_i})) + 0 \bullet \Lambda(-\mathbf{Z_i}) = 1-\Lambda(-\mathbf{Z_i})$$

Y por tanto

$$\frac{\partial E\left[\mathbf{y_i}\left|\mathbf{x_i}\right.\right]}{\partial \mathbf{X_j}} = -\frac{\partial \Lambda\left(-\mathbf{Z_i}\right)}{\partial \mathbf{X_j}}\beta_j = \Lambda\left(\mathbf{Z_i}\right)\left[1-\Lambda\left(\mathbf{Z_i}\right)\right]\beta_j$$

Noten que para el modelo Logit también se tiene que el efecto marginal de X_j no es el mismo para todos los individuos. Al igual que en el modelo Probit, una práctica común para resumir el efecto marginal para la muestra es calcular todos los efectos marginales para la muestra y calcular el promedio de este. Es decir:

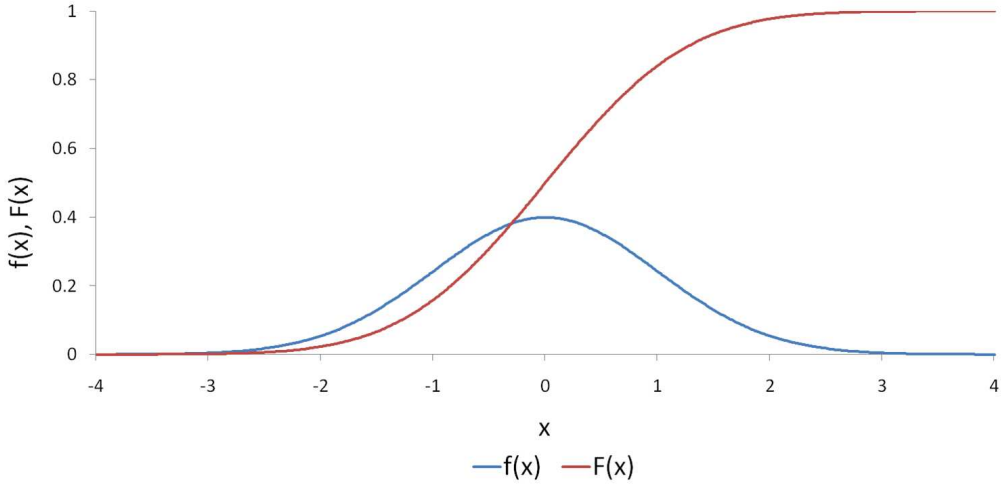
$$\frac{\sum_{i=1}^{n}\Lambda\left(Z_i\right)\left[1-\Lambda\left(Z_i\right)\right]\beta_j}{n}$$

7.4. Comentarios sobre los modelos Probit y Logit

La primera diferencia evidente entre emplear un modelo Probit o Logit es el supuesto sobre la distribución empleado. El primero supone una distribución normal mientras que el segundo implica una distribución logística. De hecho la forma de ambas distribuciones acumulativas de probabilidad es muy parecida (ver figura 7.4). Esta forma funcional tan parecida hace que exista una relación entre los coeficientes estimados con cada uno de estos modelos para una misma muestra, en especial se tiene que $\hat{\beta}_{\text{Probit}} \approx 0{,}625\hat{\beta}_{\text{Logit}}$.

Así mismo, los efectos marginales calculados por cada uno de los métodos son relativamente iguales. Por otro lado, no existe ningún método estadístico que permita decidir entre los dos modelos, pues éstos no son comparables.

Figura 7.4: Función de distribución acumulativa normal y logística ($F\left(x\right)$)



También, es importante resaltar que en este contexto no tiene mucho sentido emplear una medida como el R^2 para examinar la bondad de ajuste del modelo, pues el método de estimación no pretende minimizar la suma de los errores cuadrados. En este caso se emplea una medida de bondad de ajuste conocida como el *índice de la razón de verosimilitud*, comúnmente conocida como el LRI (por la sigla

del término en inglés Likelihood Ratio Index). Esta medida tiene similitudes con el R^2 ajustado al no tener interpretación clara, y no verse afectado por la exclusión o inclusión de variables, por otro lado el LRI se encuentra acotado entre 0 y 1; de tal manera que cuanto más grande el índice, mejor será el ajuste del modelo.

El LRI se calcula de la siguiente manera:

$$LRI = 1 - \frac{\ln L\left(\hat{\beta}\right)}{\ln L_0}$$

Donde $\ln L\left(\hat{\beta}\right)$ corresponde al valor de la función de verosimilitud en su máximo y $\ln L_0$ corresponde al valor del logaritmo de la función de verosimilitud si todos los parámetros fueran cero, excepto el intercepto. Este último valor se puede calcular de la siguiente manera:

$$\ln L_0 = n\left[P \ln\left(P\right) + \left(1 - P\right) \ln\left(1 - P\right)\right]$$

Donde P es la proporción de observaciones para las cuales la dummy es igual a uno, es decir $P = (\#obs = 1)/n$. Antes de discutir un ejemplo, es importante anotar que el LRI no es comparable entre modelos Logit y Probit, pero si permite comparar modelos Logit entre sí o modelos Probit entre sí.

En la primera sección se discutió la necesidad de emplear una variable latente (Z_i) no observable para estimar este tipo de modelos. Si bien esa variable no es observable, se puede estimar empleando los β estimados. Es decir,

$$\hat{Z}_i = \hat{\beta}^T \mathbf{x}_i^T$$

Por otro lado, el error ε_i que no es observable no puede ser estimado en este caso, pues no se cuenta con Z_i, para calcular $\hat{\varepsilon}_i = Z_i - \hat{Z}_i$.

7.5. Determinantes de la probabilidad de jugar chance

El objetivo de este ejercicio es determinar cuáles son las características que determinan la probabilidad de convertirse en jugador de apuestas permanentes (chance) en la zona comprendida por Cali, Yumbo y Jamundí.[6] Es así como partir de la información recolectada en las encuestas a los hogares, se evaluaron modelos Probit y Logit para esta zona. Las variables usadas fueron el estrato, la edad y el nivel educativo como posibles determinantes que inciden que el encuestado juegue chance. Antes de continuar definamos la variable dependiente. Sea:

$$fjuega_i = \begin{cases} 1 & \text{si el individuo } i \text{ juega chance} \\ 0 & \text{o.w} \end{cases}$$

7.5.1. Estimación del modelo Probit por el método de máxima verosimilitud

Recuerden que el modelo Probit para explicar si un jugador de juegos de azar juega chance está dado por:

$$P(fjuega = 1|\, \mathbf{x}_i) = \Phi(\beta^T \mathbf{x}_i^T)$$

[6] Estos municipios hacen parte del departamento del Valle del Cauca en Colombia.

donde $\Phi(Z) = \int_{-\infty}^{Z} \phi(v)\,dv = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z} e^{-\frac{v}{2}} dv$, es decir la función de densidad de la normal estándar; $\mathbf{X_i}$ corresponde al vector fila de las características del encuestado i para jugar, es decir:

$$\mathbf{x}_i = \left[\begin{array}{cccc} 1 & Estrato_i & Niveleducativo_i & Edad_i \end{array} \right]$$

y además,

$$\beta^T = \left[\begin{array}{cccc} \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{array} \right]$$

Por tanto: $\beta^T \mathbf{x}_i^T = \beta_1 + \beta_2 Estrato_i + \beta_3 Niveleducativo_i + \beta_4 Edad_i + \varepsilon_i$

El vector de coeficientes β puede ser estimado por medio del método de máxima verosimilitud, que equivale a resolver el siguiente problema:

$$\underset{\hat{\beta}}{Max}\, P(y\,|\mathbf{x}_i) = \underset{\hat{\beta}}{Max} \left[\prod_{i=1}^{n} \Phi\left(\hat{\beta}^T {\mathbf{x}_i}_i^T\right) \right]$$

Lo cual es equivalente a:

$$\underset{\hat{\beta}}{Max}\, l(y\,|\mathbf{x}_i) = \underset{\hat{\beta}}{Max} \left[\prod_{i=1}^{n} \ln\left[\Phi\left(\hat{\beta}^T \mathbf{X_i^T}\right)\right] \right]$$

donde $\ln\left[\Phi\left(\beta^T \mathbf{x}_i^T\right)\right]$ corresponde al logaritmo natural y $l(y\,|\mathbf{x}_i)$ es conocida como el logaritmo de la función de verosimilitud. La solución a este problema no es trivial y debe solucionarse numéricamente. Para encontrar la solución de este problema se puede emplear EasyReg. Antes de iniciar, tenga en cuenta que las variables dummy ya han sido creadas en Excel y se definen de la siguiente manera, en el caso de la educación:

$$est2_i = \left\{ \begin{array}{ll} 1 & \text{si el individuo } i \in \text{ al estrato 2} \\ 0 & o.w. \end{array} \right.$$

$$est3_i = \left\{ \begin{array}{ll} 1 & \text{si el individuo } i \in \text{ al estrato 3} \\ 0 & o.w. \end{array} \right.$$

$$est4_i = \left\{ \begin{array}{ll} 1 & \text{si el individuo } i \in \text{ al estrato 4} \\ 0 & o.w. \end{array} \right.$$

$$est5_i = \left\{ \begin{array}{ll} 1 & \text{si el individuo } i \in \text{ al estrato 5} \\ 0 & o.w. \end{array} \right.$$

$$est6_i = \left\{ \begin{array}{ll} 1 & \text{si el individuo } i \in \text{ al estrato 6} \\ 0 & o.w. \end{array} \right.$$

Y en cuanto al nivel educativo, hemos creado las variables dummy con la siguiente especificación:

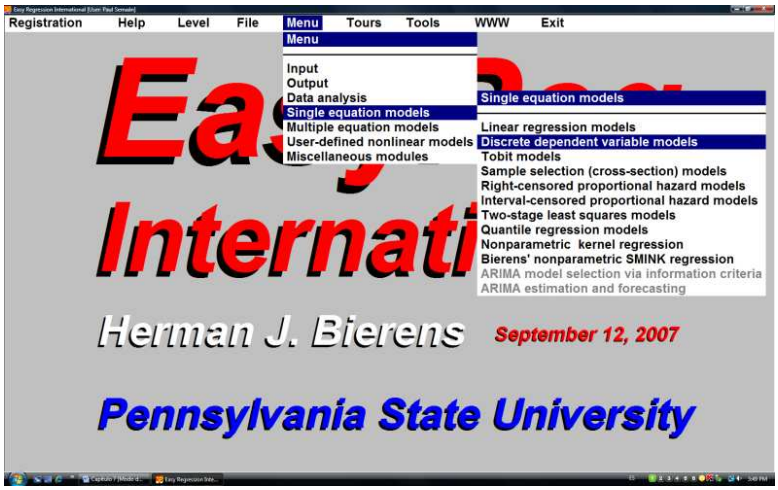
$$edu1_i = \left\{ \begin{array}{ll} 1 & \text{si el individuo } i \text{ completó el bachillerato} \\ 0 & o.w. \end{array} \right.$$

$$edu2_i = \left\{ \begin{array}{ll} 1 & \text{si el individuo } i \text{ no completó la universidad} \\ 0 & o.w. \end{array} \right.$$

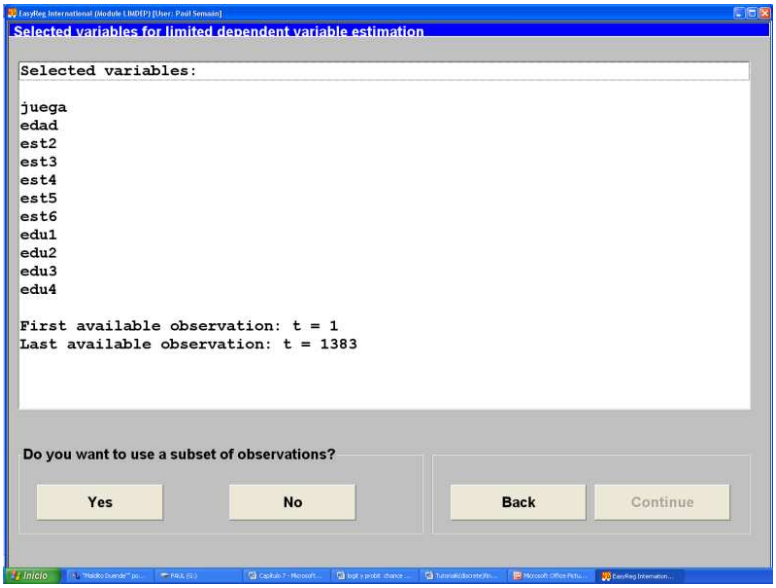
$$edu3_i = \left\{ \begin{array}{ll} 1 & \text{si el individuo } i \text{ completó la universidad} \\ 0 & o.w. \end{array} \right.$$

$$edu4_i = \begin{cases} 1 & \text{si el individuo } i \text{ tiene post} - \text{grado} \\ 0 & o.w. \end{cases}$$

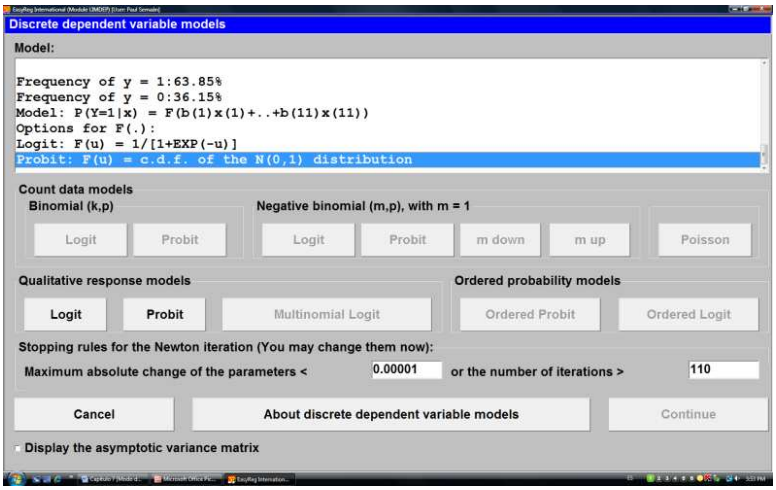
Después de haber cargado los datos "elogprob", en el menú principal haga clic en "Menu/Single equation/Discrete dependent variable models".



Ahora, escoja haciendo doble clic, las siguientes variables "juega", "edad", "est2", "est3", "est4", "est5", 2est6", "edu1", "edu2", "edu3" y "edu4". Ahora haga clic en "Selection OK". Observará la siguiente ventana.



Haga clic en "No" y en "Continue". Ahora escoja la variable dependiente, es decir "juega" y haga clic en "Continue" dos veces. En la siguiente ventana verá las demás variables que serán empleadas como variables explicativas, asegúrese que están seleccionadas todas las variables explicativas que desea y haga clic en "Selection OK". Ahora haga clic en "Continue" dos veces. Observará la siguiente ventana.



En esta ventana usted debe escoger el modelo a estimar, puede escoger un modelo Logit o Probit. Haga clic en el botón "Probit" y posteriormente en el botón "Continue". En la siguiente ventana encontrará la estimación del modelo que se reporta completamente en el recuadro 7.7.5.1. Este resultado se encuentra resumido en el cuadro 7.1

Recuadro 7.1Resultados de la Estimación del Modelo Probit

Probit model:
Dependent variable:
Y = juega
Characteristics:
juega
First observation = 1
Last observation = 1383
Number of usable observations: 1383
Minimum value: 0.0000000E+000
Maximum value: 1.0000000E+000
Sample mean: 6.3846710E-001
This variable is a zero-one dummy variable.
A Probit or Logit model is suitable.

X variables:
X(1) = edad
X(2) = est2
X(3) = est3
X(4) = est4
X(5) = est5
X(6) = est6
X(7) = edu1
X(8) = edu2
X(9) = edu3
X(10) = edu4
X(11) = 1

Frequency of y = 1:63.85 %
Frequency of y = 0:36.15 %
Model: P(Y=1|x) = F(b(1)x(1)+..+b(11)x(11))
Chosen option: F(u) = c.d.f. of N(0,1) distr. (Probit model)
Newton iteration succesfully completed after 5 iterations
Last absolute parameter change = 0.0000
Last percentage change of the likelihood = 0.0000

Maximum likelihood estimation results:

Par.	ML estimate	s.e.	t-value	[p-value]	Variable
b(1)	0.010397	0.002561	4.06	[0.00005]	edad
b(2)	-0.198876	0.127026	-1.57	[0.11743]	est2
b(3)	-0.259373	0.127395	-2.04	[0.04175]	est3
b(4)	-0.741088	0.173714	-4.27	[0.00002]	est4
b(5)	-0.967397	0.187924	-5.15	[0.00000]	est5
b(6)	-1.216866	0.372506	-3.27	[0.00109]	est6
b(7)	-0.122518	0.083478	-1.47	[0.14219]	edu1
b(8)	-0.116267	0.159816	-0.73	[0.46692]	edu2
b(9)	-0.298365	0.155743	-1.92	[0.05540]	edu3
b(10)	-1.405799	0.573293	-2.45	[0.01420]	edu4
b(11)	0.311055	0.157215	1.98	[0.04787]	1

The two-sided p-values are based on the normal approximation

Log likelihood: -8.47862227893E+002
Sample size (n):
1383
Information criteria:

Akaike: 1.242027806
Hannan-Quinn: 1.257593515 The Hausman test could not be conducted due to the following problem:
Schwarz: 1.283641771
Variance matrix of b(NLLS)-b(ML) is not positive-definite.

Así, los coeficientes estimados corresponden a:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \\ \hat{\beta}_7 \\ \hat{\beta}_8 \\ \hat{\beta}_9 \\ \hat{\beta}_{10} \\ \hat{\beta}_{11} \end{bmatrix} = \begin{bmatrix} 0{,}3110547 \\ 0{,}0103971 \\ -0{,}1988761 \\ -0{,}2593733 \\ -0{,}7410875 \\ -0{,}9673974 \\ -1{,}2168663 \\ -0{,}1225179 \\ -0{,}1162670 \\ -0{,}2983650 \\ -1{,}4057989 \end{bmatrix}$$

La correspondiente matriz de varianzas y covarianzas se puede encontrar empleando la matriz final de los resultados reportados en el recuadro 7.7.5.1.

Noten que el logaritmo de la función de máxima verosimilitud en su máximo corresponde a -847.862 = -8.47862227893E + 002. Empleando Excel, se puede observar rápidamente que el Índice de la razón de Verosimilitud (LRI) para este caso corresponde a:

$$LRI = 1 - \frac{\ln L}{\ln L_0} = 1 - \frac{-847{,}862}{-904{,}89} = 0{,}063$$

Donde $\ln L_0 = n\left[P \ln(P) + (1-P)\ln(1-P)\right]$, $n = 1383$ corresponde al número de observaciones y es la proporción de observaciones que toman el valor de uno. En el archivo "Anexologprob.xls" se presenta una hoja electrónica que efectúa este cálculo.

Para encontrar una interpretación intuitiva de los coeficientes, es importante calcular el efecto marginal promedio de cada variable. Esto se puede efectuar por medio de Excel, calculando el promedio muestral de cada uno de los efectos marginales. Esto se puede hacer evaluando para cada una de las 1383 observaciones la siguiente fórmula:[7]

$$\frac{\partial E\left[y_i \left| \beta^T \mathbf{x_i^T}\right.\right]}{\partial \mathbf{X_{j,i}}} = \varphi\left(\beta^T \mathbf{x_i^T}\right)\beta_j = \frac{1}{\sqrt{2\pi}} e^{-\frac{\left(\beta^T \mathbf{x_i^T}\right)^2}{2}} \cdot \beta_j$$

donde $i = 1, 2, ...n$ y $j = 1, 2, ...k$; después de calcular esta fórmula se calcula el promedio para cada j. En el archivo "Anexologprob.xls" se presentan los cálculos correspondientes, asegúrese que usted entiende estos cálculos y resultados.

El resultado del cálculo del efecto marginal promedio se reporta en el cuadro 7.1. Por ejemplo, encontramos que en promedio un año mas de edad implica un incremento en la probabilidad de jugar chance de 0.36 puntos porcentuales.

[7] En la página web del libro encontrará un archivo de Excel que acompaña a este capítulo y que le permite calcular el LRI.

Noten que EasyReg nos permite probar cualquier tipo de restricción lineal de la forma $R\beta = C$ por medio de la prueba de Wald. Esta opción se puede acceder por medio del menú "Options/Wald test of linear parameter restrictions". Para continuar, haga clic en la opción "Done" para regresar al menú principal.

Cuadro 7.1: Estimación del Modelo Probit y Logit

	Variable dependiente $juega_i$ Estadísticos t entre paréntesis			
	Probit EMV		Logit EMV	
	Coeficientes	Efecto Marginal	Coeficientes	Efecto Marginal
Constante	0.3110547	–	0.4996	–
	(1.98)**		(1.90)*	
edad$_i$	0.0104	0.0037	0.0173	0.0037
	(4.06)***		(4.04)***	
est2$_i$	-0.1989	-0.0699	-0.3358	-0.0716
	(-1.57)		(-1.56)	
est3$_i$	-0.2594	-0.0912	-0.4318	-0.0921
	(-2.04)**		(-2.01)**	
est4$_i$	-0.7411	-0.2605	-1.2107	-0.2583
	(-4.27)***		(-4.22)***	
est5$_i$	-0.9674	-0.3400	-1.5739	-0.3357
	(-5.15)***		(-5.06)***	
est6$_i$	-1.2169	-0.4288	-1.9925	-0.4250
	(-3.27)***		(-3.16)***	
edu1$_i$	-0.1225	-0.0464	-0.2031	-0.0433
	(-1.47)		(-1.49)	
edu2$_i$	-0.1163	-0.0466	-0.1877	-0.0400
	(-0.73)		(-0.72)	
edu3$_i$	-0.2984	-0.1109	-0.4817	-0.1028
	(-1.92)*		(-1.91)*	
edu4$_i$	-1.4058	-0.4949	-2.3258	-0.4961
	(-2.45)**		(-2.18)**	
LRI	0.06302		0.06289	
Wald	101.87***		95.54***	
No. de obs.	1383		1383	

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos Cuadrados Ordinarios

Wald: corresponde al estadístico de Wald que comprueba la significancia conjunta de todas las pendientes

EMV: Estimadores de máxima verosimilitud

7.5.2. Estimación del modelo Logit por el método de máxima verosimilitud

Recuerden que la única diferencia entre el modelo Probit y el modelo Logit corresponde al supuesto de la función de densidad que se emplea. En este caso, el modelo Logit para explicar la probabilidad de jugar chance está dada por:

$$P\left(fjuega_i = 1 \left| \mathbf{x}_i^T \right.\right) = \Lambda\left(\beta^T \mathbf{x}_i^T\right)$$

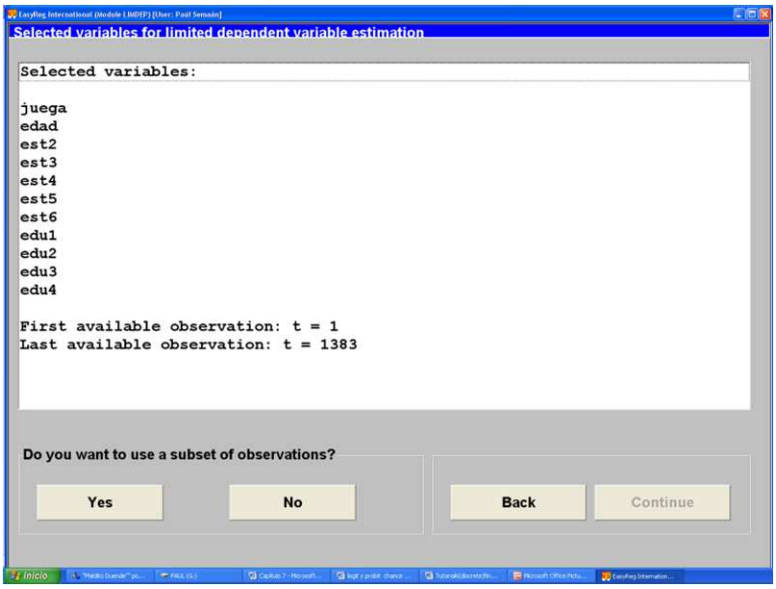
donde $\Lambda(Z) = \frac{e^Z}{1+e^Z}$, es decir la función de densidad logística. Como se discutió, el vector de coeficientes β puede ser estimado por medio del método de máxima verosimilitud, que equivale a resolver el siguiente problema:

$$\underset{\hat{\beta}}{Max}\, P\left(Y \left| X \right.\right) = \underset{\hat{\beta}}{Max} \left[\prod_{i=1}^{n} \Lambda\left(\beta^T \mathbf{x}_i{}^{\mathbf{T}}\right)\right]$$

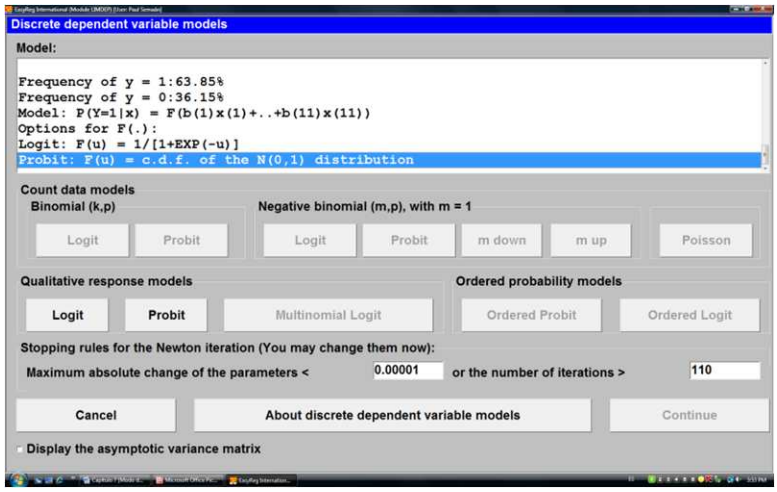
Este problema es equivalente a resolver lo siguiente:

$$\underset{\hat{\beta}}{Max}\, l\left(Y \left| X \right.\right) = \underset{\hat{\beta}}{Max} \left[\prod_{i=1}^{n} \ln\left[\Lambda\left(\beta^T \mathbf{x_i^T}\right)\right]\right]$$

donde $\ln[\bullet]$ corresponde al logaritmo natural y $l\left(Y \left| X \right.\right)$. es conocida como el logaritmo de la función de verosimilitud. EasyReg resuelve este problema automáticamente. Para encontrar la solución, haga clic en "Menu/Single equation/Discrete dependent variable models". Y escoja, haciendo doble clic, las siguientes variables "juega", "edad", "est2", "est3", "est4", "est5", "est6", "edu1", "edu2", "edu3" y "edu4". Ahora haga clic en "Selection OK". Observará la siguiente ventana.



Haga clic en "No" y en "Continue". Ahora escoja la variable dependiente, es decir "juega" y haga clic en "Continue" dos veces. En la siguiente ventana verá las demás variables que serán empleadas como variables explicativas, asegúrese que están seleccionadas todas las variables explicativas que desea y haga clic en "Selection OK". Ahora haga clic en "Continue" dos veces. Observará la siguiente ventana.



En esta ventana haga clic en el botón "Logit" y posteriormente en el botón "Continue". En la siguiente ventana encontrará la estimación del modelo que se reporta completamente en el recuadro 7.2. Este resultado se encuentra resumido en el cuadro 7.1.

Recuadro 7.2 Resultados de la Estimación del Modelo Probit

Logit model:
Dependent variable:
Y = juega
Characteristics:
juega
First observation = 1
Last observation = 1383
Number of usable observations: 1383
Minimum value: 0.0000000E+000
Maximum value: 1.0000000E+000
Sample mean: 6.3846710E-001
This variable is a zero-one dummy variable.
A Probit or Logit model is suitable.

X variables:
X(1) = edad
X(2) = est2
X(3) = est3
X(4) = est4
X(5) = est5
X(6) = est6
X(7) = edu1
X(8) = edu2
X(9) = edu3
X(10) = edu4
X(11) = 1

Frequency of y = 1:63.85 %
Frequency of y = 0:36.15 %
Model: P(Y=1|x) = F(b(1)x(1)+..+b(11)x(11))
Chosen option: F(u) = 1/[1+EXP(-u)] (Logit model)

Newton iteration succesfully completed after 6 iterations
Last absolute parameter change = 0.0000
Last percentage change of the likelihood = 0.0000

Maximum likelihood estimation results:

Par.	ML estimate	s.e.	t-value	[p-value]	Variable
b(1)	0.017276	0.004272	4.04	[0.00005]	edad
b(2)	-0.335755	0.214745	-1.56	[0.11793]	est2
b(3)	-0.431835	0.214973	-2.01	[0.04456]	est3
b(4)	-1.210719	0.287094	-4.22	[0.00002]	est4
b(5)	-1.573915	0.310865	-5.06	[0.00000]	est5
b(6)	-1.992522	0.631231	-3.16	[0.00160]	est6
b(7)	-0.203087	0.136402	-1.49	[0.13652]	edu1
b(8)	-0.187661	0.259462	-0.72	[0.46951]	edu2
b(9)	-0.481720	0.251611	-1.91	[0.05555]	edu3
b(10)	-2.325827	1.069145	-2.18	[0.02960]	edu4
b(11)	0.499578	0.263463	1.90	[0.05793]	1

The two-sided p-values are based on the normal approximation

Log likelihood: -8.47979855309E+002
Sample size (n):
1383
Information criteria:
Akaike: 1.242197911
Hannan-Quinn: 1.257763619
Schwarz: 1.283811876
The Hausman test could not be conducted due to the following problem:
Variance matrix of b(NLLS)-b(ML) is not positive-definite.

En este caso tenemos que los coeficientes estimados para el modelo Logit son:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \\ \hat{\beta}_5 \\ \hat{\beta}_6 \\ \hat{\beta}_7 \\ \hat{\beta}_8 \\ \hat{\beta}_9 \\ \hat{\beta}_{10} \\ \hat{\beta}_{11} \end{bmatrix} = \begin{bmatrix} 0{,}4995778 \\ 0{,}0172758 \\ -0{,}3357550 \\ -0{,}4318348 \\ -1{,}2107193 \\ -1{,}5739152 \\ -1{,}9925222 \\ -0{,}2030874 \\ -0{,}1876610 \\ -0{,}4817196 \\ -2{,}3258271 \end{bmatrix}$$

Noten que los coeficientes estimados para el modelo Probit y el modelo Logit conservan el mismo signo y como lo sugiere Amemiya (1981) los estimadores de máxima verosimilitud para los modelo Probit y Logit siguen aproximadamente la siguiente relación:

$$\hat{\beta}_{\text{Probit}} \approx 0{,}625\hat{\beta}_{\text{Logit}}$$

En este caso, ustedes pueden comprobar que esta relación se mantiene aproximadamente. Recuerde que esta relación se da gracias a la similitud entre las dos funciones de densidad (ver figura 7.4).

Para este modelo Logit estimado tenemos que el logaritmo de la función de máxima verosimilitud en su máximo corresponde a -847.979 = -8.47979855309E + 002. Como se mencionó anteriormente,

por medio de Excel se puede encontrar rápidamente que el Índice de la razón de Verosimilitud (LRI) para este caso corresponde a:

$$LRI = 1 - \frac{\ln L}{\ln L_0} = 1 - \frac{-847{,}979}{-904{,}89} = 0{,}06289$$

donde $\ln L_0 = n\left[P \ln(P) + (1-P)\ln(1-P)\right]$, $n = 1383$ corresponde al número de observaciones y $P = 0{,}638$ es la proporción de observaciones que toman el valor de uno.

Para encontrar una interpretación intuitiva de los coeficientes, es importante calcular el efecto marginal promedio de cada variable. Esto se puede efectuar por medio de Excel, calculando el promedio muestral de cada uno de los efectos marginales. Esto se puede hacer evaluando para cada una de las 1383 observaciones la siguiente formula

Así como en el modelo Probit, para encontrar una interpretación intuitiva de los coeficientes, es importante calcular el efecto marginal promedio de cada variable. Esto se puede efectuar por medio de Excel, calculando el promedio muestral de cada uno de los efectos marginales. Esto se puede hacer evaluando para cada una de las 1383 observaciones la siguiente fórmula:

Así como en el modelo Probit, para encontrar una interpretación intuitiva de los coeficientes, es importante calcular el efecto marginal promedio de cada variable. Esto se puede efectuar por medio de Excel, calculando el promedio muestral de cada uno de los efectos marginales. Esto se puede hacer evaluando para cada una de las 1383 observaciones la siguiente fórmula:

$$\frac{\partial E\left[y_{\mathbf{i}} \left| \beta^T \mathbf{x_i^T}\right.\right]}{\partial \mathbf{x_{j,i}}} = \Lambda\left(\beta^T \mathbf{x_i^T}\right)\left[1 - \Lambda\left(\beta^T \mathbf{x_i^T}\right)\right]\beta_j = \left(\frac{e^{\beta^T \mathbf{x_i^T}}}{1 + e^{\beta^T \mathbf{x_i^T}}}\right)\left[1 - \left(\frac{e^{\beta^T \mathbf{x_i^T}}}{1 + e^{\beta^T \mathbf{x_i^T}}}\right)\right] \cdot \beta_j$$

donde $i = 1, 2, ...n$ y $j = 1, 2, ...k$; después de calcular esta fórmula se calcula el promedio para cada j.

Los efectos marginales estimados se reportan en el cuadro 7.1. Noten que estos efectos marginales del modelo Logit son muy similares a los efectos marginales encontrados para el modelo Probit estimado.

Noten que así como para el caso del modelo Probit, EasyReg nos permite probar cualquier tipo de restricción lineal de la forma $R\beta = C$ por medio de la prueba de Wald. Esta opción se puede acceder por medio del menú "Options/Wald test of linear parameter restrictions". Haga clic en la opción "Done" para regresar al menú principal.

7.6. Ejercicios

Trece aspirantes a una beca estatal para subsidiar programas de doctorado obtuvieron diferentes resultados en sus pruebas cuantitativas y verbales. Seis de los trece aspirantes fueron admitidos al programa. Con base en los datos del archivo logprob.xls conteste las siguientes preguntas:

1. Estime el modelo MLP para estimar la probabilidad de admisión al programa con base en los puntajes cuantitativos y cualitativos de la prueba.

2. ¿Considera usted que el modelo estimado en el punto 1, es un modelo satisfactorio? Sustente suficientemente su respuesta.

3. Estime el modelo Logit y reporte sus resultados en una tabla, no olvide mostrar en ésta cuál es la razón de máxima verosimilitud y el logaritmo de la función de máxima verosimilitud. Comente sobre la significancia conjunta e individual de los coeficientes.

4. Estime el modelo Probit y reporte sus resultados en una tabla, no olvide mostrar en ésta cuál es la razón de máxima verosimilitud y el logaritmo de la función de máxima verosimilitud. Comente sobre la significancia conjunta e individual de los coeficientes.

5. Encuentre el efecto marginal de los resultados en las pruebas cuantitativas y verbales para ambos modelos para cada uno de los trece individuo, a su vez determine el efecto marginal promedio.

7.7. Apéndice

Apéndice 7.1 Demostración de la presencia de Heteroscedasticidad en un Modelo de Probabilidad Lineal

El Modelo de Probabilidad Lineal (MPL) corresponde a un modelo lineal cuya variable dependiente es una dummy. Supongamos que:

$$P\left[\mathbf{y}=1\right]=F\left(\beta^{T}\mathbf{X_i^T}\right)$$

$$P\left[\mathbf{y}=0\right]=1-F\left(\beta^{T}\mathbf{X_i^T}\right)$$

Donde, $\mathbf{X_i}=\left[\begin{array}{cccccc}1 & \mathbf{X_{2,i}} & \mathbf{X_{3,i}} & \cdots & \mathbf{X_{k,i}}\end{array}\right]_{1\times k}$ corresponde a un vector fila con todas las características del individuo i, y $\beta^{T}=\left[\begin{array}{cccccc}\beta_1 & \beta_2 & \beta_3 & \cdots & \beta_k\end{array}\right]$. Por tanto, el valor esperado de la variable dependiente será:

$$E\left[\mathbf{y_i}\left|\beta^{T}\mathbf{X_i^T}\right.\right]=1\bullet F\left(\beta^{T}\mathbf{X_i^T}\right)+0\bullet\left(1-F\left(\beta^{T}\mathbf{X_i^T}\right)\right)$$

$$E\left[\mathbf{y_i}\left|\beta^{T}\mathbf{X_i^T}\right.\right]=F\left(\beta^{T}\mathbf{X_i^T}\right)$$

El MPL implica suponer que

$$F\left(\beta^{T}\mathbf{X_i^T}\right)=\beta^{T}\mathbf{X_i^T}=\beta_1+\beta_2X_{2,i}+\beta_3X_{3,i}+...+\beta_kX_{k,i}$$

$$F\left(\beta^{T}\mathbf{X_i^T}\right)=\beta^{T}\mathbf{X_i^T}$$

Así, el modelo que describe el comportamiento de la variable dummy será

$$\mathbf{y_i}=E\left[\mathbf{y_i}\left|\beta^{T}\mathbf{X_i^T}\right.\right]+\varepsilon_i$$

$$\mathbf{y_i}=F\left(\beta^{T}\mathbf{X_i^T}\right)+\varepsilon_i$$

$$Y_i=\beta_1+\beta_2X_{2,i}+\beta_3X_{3,i}+...+\beta_kX_{k,i}+\varepsilon_i$$

Este modelo lo podemos escribir de forma más compacta de la siguiente manera:

$$\mathbf{y_i}=\beta^{T}\mathbf{X_i^T}+\varepsilon_i$$

Supongamos que los errores tienen media cero y no están relacionados. Es decir, $E\left[\varepsilon_i\right]=0$ y $E\left[\varepsilon_i,\varepsilon_j\right]=0\ \forall\ i\neq j$. Además,

$$Y_i=\begin{cases}1 & \text{con probabilidad F}\\ 0 & \text{con probabilidad } 1-\text{F}\end{cases}$$

Por simplicidad escribiremos $F=F\left(\beta^{T}\mathbf{X_i^T}\right)=\beta^{T}\mathbf{X_i^T}$. Como $E\left[\varepsilon_i\right]=0$, entonces $Var\left[\varepsilon_i\right]=E\left[\varepsilon_i^2\right]-\left[E\left[\varepsilon_i\right]\right]^2=E\left[\varepsilon_i^2\right]-0=E\left[\varepsilon_i^2\right]$. Así, debemos determinar a que es igual el valor esperado

del error al cuadrado para determinar si éste es constante o no, para saber si es homoscedástico o heteroscedástico.

Noten que,

$$\mathbf{y_i} = \beta^T \mathbf{X_i^T} + \varepsilon_i = \begin{cases} 0 & \text{con probabilidad } 1 - \text{F} \\ 1 & \text{con probabilidad F} \end{cases}$$

Esto implica que el término de error tomará únicamente dos posibles valores dado un vector de características $\mathbf{X_i^T}$. Es decir,

$$\varepsilon_i = \begin{cases} -\beta^T \mathbf{X_i^T} & \text{con probabilidad } 1 - \text{F} \\ 1 - \beta^T \mathbf{X_i^T} & \text{con probabilidad F} \end{cases}$$

Dado que deseamos encontrar la varianza del error y esta corresponde al valor esperado del cuadrado del error ($Var\left[\varepsilon_i\right] = E\left[\varepsilon_i^2\right]$), entonces tenemos que tener la expresión del cuadrado del error. Es decir,

$$\varepsilon_i^2 = \begin{cases} \left[-\beta^T \mathbf{x}_i^T\right]^2 & \text{con probabilidad } 1 - \text{F} \\ \left[1 - \beta^T \mathbf{x}_i^T\right]^2 & \text{con probabilidad F} \end{cases}$$

O lo que es lo mismo,

$$\varepsilon_i^2 = \begin{cases} \beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T} & \text{con probabilidad } 1 - \text{F} \\ 1 - 2\beta^T \mathbf{X_i^T} + \beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T} & \text{con probabilidad F} \end{cases}$$

Como se trata de una variable discreta, el valor esperado del error al cuadrado corresponderá a:

$$E\left[\varepsilon_i^2\right] = \left[\beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right][1 - F] + \left[1 - 2\beta^T \mathbf{X_i^T} + \beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right] F$$

$$E\left[\varepsilon_i^2\right] = \left[\beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right] - \left[\beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right] F + F - 2\left[\beta^T \mathbf{X_i^T}\right] F + \left[\beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right] F$$

$$E\left[\varepsilon_i^2\right] = \left[\beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right] + F - 2\left[\beta^T \mathbf{X_i^T}\right] F$$

Como $F = \beta^T \mathbf{X_i^T}$, tenemos que:

$$E\left[\varepsilon_i^2\right] = \left[\beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right] + F - 2\left[\beta^T \mathbf{X_i^T}\right] \beta^T \mathbf{X_i^T}$$

$$E\left[\varepsilon_i^2\right] = F - \left[\beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right] = \beta^T \mathbf{X_i^T} - \left[\beta^T \mathbf{X_i^T} \beta^T \mathbf{X_i^T}\right]$$

$$E\left[\varepsilon_i^2\right] = \beta^T \mathbf{X_i^T}\left[1 - \beta^T \mathbf{X_i^T}\right]$$

Es decir,

$$Var\left[\varepsilon_i\right] = E\left[\varepsilon_i^2\right] = \beta^T \mathbf{X_i^T}\left[1 - \beta^T \mathbf{X_i^T}\right]$$

$$Var\left[\varepsilon_i\right] = E\left[\mathbf{y_i} \left| \mathbf{X_i} \right.\right]\left[1 - E\left[\mathbf{y_i} \left| \mathbf{X_i} \right.\right]\right]$$

Esto implica que la varianza del error en un modelo de probabilidad lineal depende de cada observación. En otras palabras, por construcción la varianza del término de error en un modelo de probabilidad lineal no será constante. Por tanto siempre que la variable dependiente sea una variable dicotómica un modelo lineal presentará el problema de heteroscedasticidad.

CAPÍTULO 8

Ecuaciones simultáneas

Objetivos del capítulo

Al finalizar este capítulo, el lector estará en capacidad de:

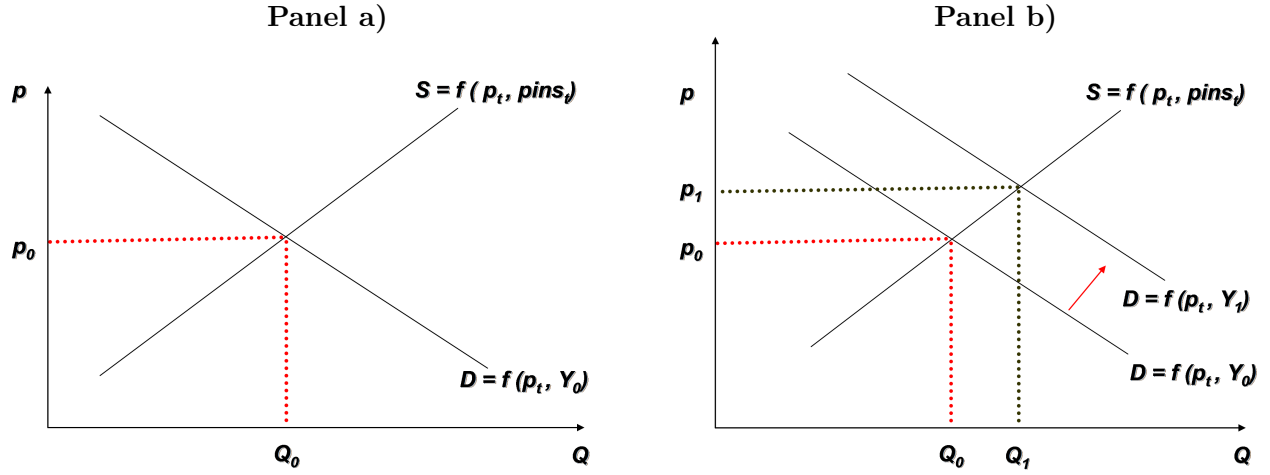
- Estimar sistemas de ecuaciones simultáneas de la forma estructural en EasyReg mediante el método de mínimos cuadrados en dos etapas .
- Efectuar la prueba de simultaneidad de Hausman mediante EasyReg.

8.1. Introducción

Hasta ahora hemos considerado modelos estadísticos lineales de una sola ecuación. Ahora concentraremos nuestra atención en modelos que tengan la intención de explicar más de una variable. Es decir, en los que más de una variable son determinadas simultáneamente.[1]

En la figura 8.1, podemos observar el equilibrio en un mercado cualquiera, en el cual la oferta depende del precio del bien y de los precios de los insumos ($pins_t$) y por su parte la demanda también depende del precio del bien (p_t) y del ingreso (Y_t).

Figura 8.1: Oferta y demanda de un Mercado



El comportamiento del mercado, que se presenta en la figura 8.1, es posible representarlo mediante las siguientes ecuaciones de oferta y demanda:

$$Q_t^d = \alpha_1 + \alpha_2 p_t + \alpha_3 Y_t + \varepsilon_t^d \tag{8.1}$$

$$Q_t^S = \beta_1 + \beta_2 p_t + \beta_3 pins_t + \upsilon_t^s \tag{8.2}$$

$$Q_t^d = Q_t^S = Q_t \tag{8.3}$$

Ahora, supongamos que estamos estudiando un bien normal ($\frac{\partial Q_t^d}{\partial p_t} = \alpha_3 > 0$) y que el ingreso aumenta. Este aumento implica un desplazamiento de la curva de demanda (a la derecha) y con esta nueva demanda se determina simultáneamente el nuevo precio y las cantidades de equilibrio. (Ver panel b) de la figura 8.1. Para un economista es intuitivo, que tanto el precio como las cantidades transadas en el mercado se determinan al interior del mercado (es decir, del sistema de ecuaciones), por eso p_t y Q_t se consideran variables endógenas. Por otro lado, el ingreso (Y_t) y el precio de los insumos ($pins_t$) son cantidades que se determinan por fuera de este mercado y por tanto para este sistema de ecuaciones de oferta y demanda se considerarán variables exógenas. En otras palabras, Y_t y $pins_t$ son variables que proveen información al sistema conformado por 8.1 8.2 8.3. Es importante mencionar que las ecuaciones 8.1 y 8.2 son denominadas ecuaciones estructurales y sus coeficientes se conocen como parámetros estructurales. Por otro lado 8.3 es una identidad que no posee parámetros a ser estimados.

[1] El estudio de Ecuaciones simultaneas fue iniciado por la "Cowless Commission" de la Universidad de Chicago en los años 40. Estos modelos fueron muy usados en las décadas del 50, 60 y 70 en la macroeconomía, sin embargo hoy en día no son tan empleados en esta área, aunque aún son útiles en otros campos.

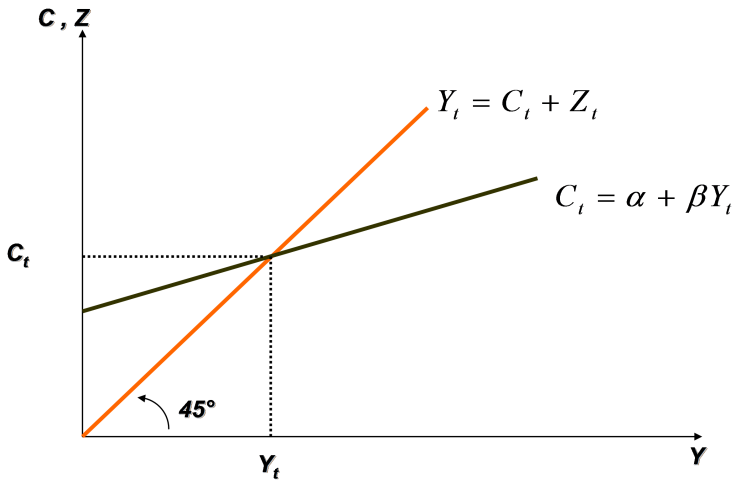
Veamos otro ejemplo. En la figura 8.2 se presenta un modelo de determinación del ingreso (Y_t), cuya especificación es:

$$C_t = \alpha + \beta Y_t + u_t \tag{8.4}$$

$$Y_t = C_t + Z_t \tag{8.5}$$

donde C_t es el consumo y Z_t representa el gasto no asociado al consumo (inversión, gasto público, etc.). En este caso, es intuitivo para un economista que C_t y Y_t son variables endógenas, mientras que Z_t es una variable exógena.

Figura 8.2: Modelo de determinación del ingreso



De los dos ejemplos anteriores, se puede observar que hay una diferencia radical con el tipo de modelos que se han estudiado hasta el momento. En las ecuaciones 8.1, 8.2 y 8.4 se tienen variables endógenas como variables explicativas. Para entender las consecuencias de tener variables endógenas al lado derecho del igual, a continuación analizaremos si los coeficientes estimados por MCO siguen siendo insesgados.

8.2. Sesgo de los estimadores MCO: Un ejemplo

Retomando el ejemplo del mercado 8.1 8.2 8.3, concentrémonos en la demanda:

$$Q_t^d = \alpha_1 + \alpha_2 p_t + \alpha_3 Y_t + \varepsilon_t^d$$

Dado que p_t es una variable endógena de este sistema, éste estará determinado por la interacción de la oferta y de la demanda y por tanto dependerá del término aleatorio de error tanto de la demanda (ε_t^d) como de la oferta (ν_t^s). Así, p_t será una variable estocástica que está actuando como variable explicativa y además tendrá una correlación con el error de esta ecuación. Esto implica violar uno de los supuestos del teorema de Gauss-Markow.

Ahora regresemos al segundo ejemplo, en el que se presenta un modelo de determinación del ingreso con la siguiente especificación

$$C_t = \alpha + \beta Y_t + u_t$$

$$Y_t = C_t + Z_t$$

Además supongamos:

$$E[u] = 0 \quad \forall t \quad \text{y} \quad E\left[u^T u\right] = \sigma^2 I_n$$

Nuestro objetivo es encontrar claramente un buen estimador de los parámetros de la función de consumo. Sean los EMCO unos buenos estimadores.

Está bien documentado que los EMCO serán MELI aún si las variables explicativas son aleatorias, siempre y cuando los regresores estocásticos sean independientes del error de la regresión (ver por ejemplo el apéndice 8.1).

En este caso podemos mostrar muy rápidamente que hay una relación lineal entre la variable explicativa Y_t y el término aleatorio de eror u_t. Por ejemplo, reemplazando:

$$C_t = \alpha + \beta Y_t + u_t$$

en

$$Y_t = C_t + Z_t$$

obtenemos:

$$Y_t = (\alpha + \beta Y_t + u_t) + Z_t$$

$$Y_t - \beta Y_t = \alpha + u_t + Z_t$$

Entonces, tenemos que

$$Y_t = \frac{\alpha}{1-\beta} + \frac{Z_t}{1-\beta} + \frac{u_t}{1-\beta}$$

Ahora veamos si en efecto existe o no relación lineal entre Y_t y u_t

$$Cov[u_t, Y_t] = E[u_t Y_t] - \{E[u_t] E[Y_t]\} = E[u_t Y_t]$$

$$Cov[u_t, Y_t] == E\left[u_t \left[\frac{\alpha}{1-\beta} + \frac{Z_t}{1-\beta} + \frac{u_t}{1-\beta}\right]\right]$$

$$Cov[u_t, Y_t] = E\left[u_t \frac{\alpha}{1-\beta} + u_t \frac{Z_t}{1-\beta} + \frac{u_t^2}{1-\beta}\right]$$

$$Cov[u_t, Y_t] = E\left[\frac{u_t^2}{1-\beta}\right] = \frac{1}{1-\beta} E\left[u_t^2\right]$$

Finalmente, tenemos que

$$Cov[u_t, Y_t] = \frac{1}{1-\beta}\sigma^2 \neq 0$$

Entonces, el término de error y la variable explicativa en el modelo que representa la función de consumo están correlacionadas. Por lo tanto como en el caso de variables explicativas medidas con error los estimadores MCO serán sesgados e inconsistentes.[2]

[2]En el apéndice 8.1 se presenta una demostración.

8.3. Forma reducida del sistema

Una opción para dar solución al problema que una variable explicativa endógena esté correlacionada con el error, es escribir el sistema de ecuaciones de tal forma que ninguna ecuación tenga variables endógenas como variables explicativas. Esto se conoce como la forma reducida del sistema. Para entender cómo funciona esta técnica, regresemos al primer ejemplo (figura 8.1).

$$Q_t^d = \alpha_1 + \alpha_2 p_t + \alpha_3 Y_t + \varepsilon_t^d$$
$$Q_t^S = \beta_1 + \beta_2 p_t + \beta_3 pins_t + \upsilon_t^s$$
$$Q_t^d = Q_t^S = Q_t$$

Resolviendo el sistema de ecuaciones para p_t tenemos lo siguiente:

$$\alpha_1 + \alpha_2 p_t + \alpha_3 Y_t + \varepsilon_t^d = Q_t^d = Q_t^S = \beta_1 + \beta_2 p_t + \beta_3 pins_t + \upsilon_t^s$$
$$\alpha_1 + \alpha_2 p_t + \alpha_3 Y_t + \varepsilon_t^d = \beta_1 + \beta_2 p_t + \beta_3 pins_t + \upsilon_t^s$$

Despejando p_t tenemos que:

$$\alpha_2 p_t - \beta_2 p_t = (\beta_1 - \alpha_1) - \alpha_3 Y_t + \beta_3 pins_t + \upsilon_t^s - \varepsilon_t^d$$
$$(\alpha_2 - \beta_2)\, p_t = (\beta_1 - \alpha_1) - \alpha_3 Y_t + \beta_3 pins_t + \upsilon_t^s - \varepsilon_t^d$$
$$p_t = \frac{1}{(\alpha_2 - \beta_2)} \left[(\beta_1 - \alpha_1) - \alpha_3 Y_t + \beta_3 pins_t + \upsilon_t^s - \varepsilon_t^d \right]$$

Finalmente p_t se encuentra en función de variables exógenas, es decir:

$$p_t = \frac{(\beta_1 - \alpha_1)}{(\alpha_2 - \beta_2)} + \left(\frac{-\alpha_3}{(\alpha_2 - \beta_2)} \right) Y_t + \left(\frac{\beta_3}{(\alpha_2 - \beta_2)} \right) pins_t + \frac{1}{(\alpha_2 - \beta_2)} \left(\upsilon_t^s - \varepsilon_t^d \right) \tag{8.6}$$

Si reemplazamos p_t de la ecuación de demanda:

$$Q_t = Q_t^d = \alpha_1 + \alpha_2 p_t + \alpha_3 Y_t + \varepsilon_t^d$$

Obtenemos:

$$\begin{aligned} Q_t =& \alpha_1 + \alpha_2 \left[\frac{(\beta_1 - \alpha_1)}{(\alpha_2 - \beta_2)} + \left(\frac{-\alpha_3}{(\alpha_2 - \beta_2)} \right) Y_t + \left(\frac{\beta_3}{(\alpha_2 - \beta_2)} \right) pins_t \right. \\ &+ \frac{1}{(\alpha_2 - \beta_2)} \left(\upsilon_t^s - \varepsilon_t^d \right) + \alpha_3 Y_t + \varepsilon_t^d \end{aligned}$$

Y además tenemos que

$$\begin{aligned} Q_t =& \left[\alpha_1 + \alpha_2 \frac{(\beta_1 - \alpha_1)}{(\alpha_2 - \beta_2)} \right] + \left(\frac{-\alpha_2 \alpha_3}{(\alpha_2 - \beta_2)} + \alpha_3 \right) Y_t \\ &+ \left(\frac{\alpha_2 \beta_3}{(\alpha_2 - \beta_2)} \right) pins_t + \left[\frac{\alpha_2}{(\alpha_2 - \beta_2)} \upsilon_t^s + \frac{1 - \alpha_2}{(\alpha_2 - \beta_2)} \varepsilon_t^d \right] \end{aligned} \tag{8.7}$$

Reparametrizando 8.6 y 8.7 obtenemos:[3]

$$p_t = \pi_{11} + \pi_{12} Y_t + \pi_{13} pins_t + \mu_{1t}$$

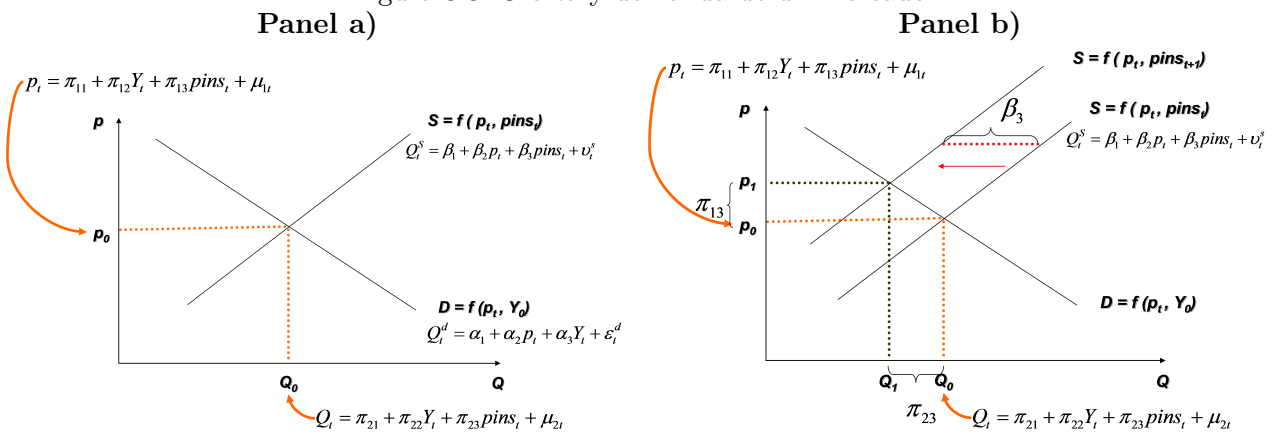
[3]Por ejemplo $\pi_{11} = \alpha_1 + \alpha_2 \frac{(\beta_1 - \alpha_1)}{(\alpha_2 - \beta_2)}$ y $\pi_{23} = \frac{(\alpha_2 \beta_3)}{(\alpha_2 - \beta_2)}$

$$Q_t = \pi_{21} + \pi_{22} Y_t + \pi_{23} pins_t + \mu_{2t}$$

Donde las variables endógenas (Q_t y p_t), están en función de variables exógenas (Y_t y $pins_t$). Es importante tener en cuenta que a estas ecuaciones se les conoce como ecuaciones reducidas y a los coeficientes como parámetros de la forma reducida.

En la figura 8.3 podemos observar la interpretación gráfica de la forma reducida. Como es posible apreciar en este ejemplo, las ecuaciones de la forma reducida, ya no explican las cantidades ofrecidas y demandadas en el mercado, tal como lo hacían las ecuaciones de la forma estructural. Por el contrario estas explican las cantidades y precios de equilibrio. En otras palabras, los coeficientes de la forma reducida no representan cambios en las cantidades ofrecidas ni demandadas sino en las cantidades y en los precios de equilibrio.

Figura 8.3: Oferta y demanda de un Mercado



Para entender la diferencia de la forma estructural y la forma reducida, supongamos que hay un aumento de una unidad en el precio del insumo (ver panel b) de la figura 8.3. El aumento en el precio de los insumos (*pins*) provoca un desplazamiento en la oferta (función de oferta) de β_3 unidades (ver panel b)), ceteris paribus. Noten que la función de demanda se mantendrá igual. El desplazamiento de la curva de oferta implica un cambio de π_{13} en el precio de equilibrio y de π_{23} en las cantidades demandadas.

Con la forma reducida, el problema de correlación entre el término de error y las variables independientes ha sido solucionado, pero recuerden que nosotros queríamos encontrar los coeficientes de la forma estructural. Es decir, lo que deseamos es conocer el valor de los coeficientes (α_1, α_2, α_3) y (β_1, β_2, β_3).

8.4. El problema de identificación

El problema de calcular los parámetros estructurales a partir de la forma reducida del sistema de ecuaciones, se conoce como problema de identificación. Diremos que una ecuación está perfectamente identificada si existe una forma unívoca de encontrar los coeficientes estructurales a partir de los coeficientes en forma reducida, es decir, que sólo hay una combinación de coeficientes estructurales que producen los coeficientes reducidos. Una ecuación está sub-identificada si no existe forma alguna de hallar los coeficientes estructurales a partir de los coeficientes en forma reducida. Finalmente, una

ecuación estará sobre-identificada si hay más de una forma de encontrar los coeficientes estructurales a partir de los coeficientes en forma reducida.

Para saber si se pueden encontrar los parámetros estructurales a partir de la forma reducida, recurrimos a la condición de orden, la cual es una condición necesaria pero no suficiente.

Sean k_j y g_j el número de variables exógenas excluidas y el número de variables endógenas incluidas en la ecuación estructural respectivamente. La condición de orden para la identificación de los parámetros de la forma estructural es $k_i \geq g_i - 1$.

Recuadro 8.1 Condición de orden

Condición	**Identificación**
$k_i = g_i - 1$	Identificación perfecta (probablemente)
$k_i > g_i - 1$	Sobre-identificación (probablemente)
$k_i < g_i - 1$	Sub-identificación (con seguridad)

Para comprender las implicaciones del problema de identificación, regresemos a nuestro ejemplo de oferta y demanda. En este caso simplificaremos su estructura de la siguiente manera:

$$Q_t^d = \alpha_1 + \alpha_2 p_t + \varepsilon_t^d \tag{8.8}$$

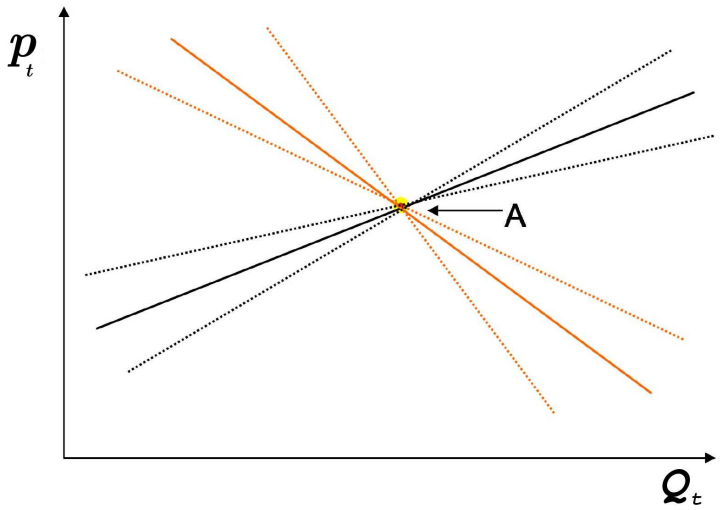
$$Q_t^S = \beta_1 + \beta_2 p_t + \upsilon_t^s \tag{8.9}$$

$$Q_t^d = Q_t^S = Q_t \tag{8.10}$$

Como podemos apreciar, estamos partiendo del supuesto de que el mercado se halla en equilibrio y por lo tanto las cantidades ofrecidas (ecuación 8.9) y las demandadas (ecuación 8.8) son iguales. Además es claro que en este ejemplo no existen variables predeterminadas o exógenas, siendo los valores de p_t y Q_t^S los únicos datos disponibles y por lo tanto siendo ambas ecuaciones sub-identificadas.

En la figura 8.4 podemos ver que el punto A representa el equilibrio en el mercado, sin embargo no podemos estimar las curvas de oferta y demanda con tan solo estos datos, debido a que en este caso no hay una forma de obtener los valores de los parámetros estructurales (intercepto y pendientes) a partir de las ecuaciones en forma reducida. En este caso, existe una cantidad infinita de curvas con diferentes pendientes e interceptos que pasan por dicho punto.

Figura 8.4: Equilibrio del mercado



Ahora bien, si tuviéramos más información, entonces podríamos identificar la función de oferta y/o demanda. Consideremos el siguiente sistema de ecuaciones:

$$Q_t^d = \alpha_1 + \alpha_2 p_t + \alpha_3 Y_t + \varepsilon_t^d \tag{8.11}$$

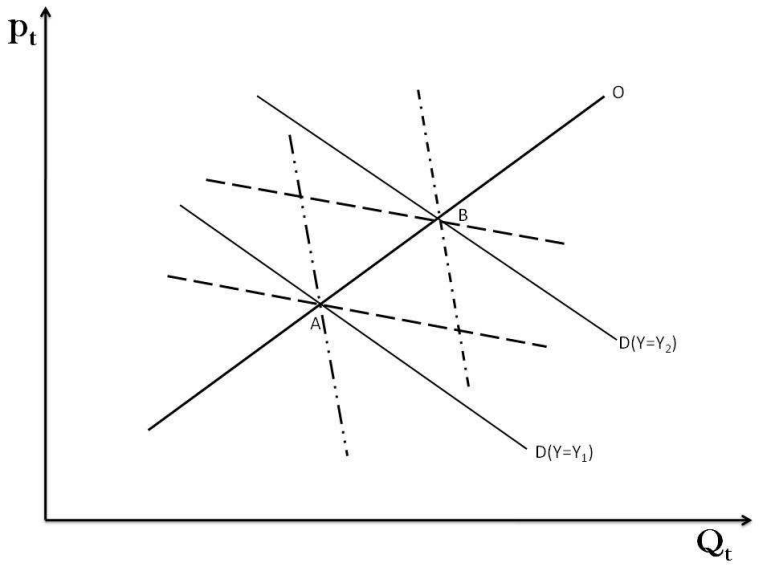
$$Q_t^S = \beta_1 + \beta_2 p_t + \upsilon_t^s \tag{8.12}$$

$$Q_t^d = Q_t^S = Q_t \tag{8.13}$$

En la ecuación 8.11, podemos apreciar que la demanda depende del ingreso, que en este caso es una variable exógena. Este hecho imposibilita que podamos trazar un conjunto de curvas de demanda y oferta infinito, ya que al variar el ingreso en el tiempo esto implicará necesariamente una cambio en la demanda y no en la oferta. En la figura 8.4 podemos apreciar que la curva de oferta se puede identificar del movimiento de la demanda sobre la oferta. Es importante reconocer que los puntos A y B representan equilibrios. Es decir, puntos en los cuales la oferta se cruza con la demanda. De acuerdo con este sistema de ecuaciones, cuando observamos un nuevo equilibrio, tiene que ser cierto que fue fruto del desplazamiento de la demanda (la cual se mueve por un cambio en el ingreso). Así, por los puntos A y B sólo puede pasar una función de oferta, pero muchas funciones de demanda. (Ver figura 8.4) Por eso se podrá recuperar la ubicación de la oferta pero no de la demanda.

La curva de oferta es identificada gracias a que los valores de sus parámetros se pueden deducir de la forma reducida y es precisamente la inclusión de una variable exógena, como el ingreso en la ecuación de demanda, lo que determina la identificación de la curva de oferta.

Figura 8.5: Identificación de la curva de oferta



Empleando la condición de orden podemos llegar a la misma conclusión a la que hemos llegado intuitivamente. Para el sistema de ecuaciones conformado por 8.11 y 8.12, tenemos que Q_t y p_t son variables endógenas y Y_t es una variable exógena. Así para la primera ecuación del sistema (ecuación 8.11) tenemos que k_1=0 y g_1=2. De esta manera tenemos que $k_1 < g_1 - 1$, es decir la ecuación de demanda estará con seguridad subidentificada. Por otro lado, para la segunda ecuación 8.12, tenemos que k_2=1 y g_2=2. Esto implica que $k_2 = g_2 - 1$ y por lo tanto la ecuación de oferta está probablemente perfectamente identificada. Esto confirma lo observado en la figura (8.4).

8.5. Estimación por mínimos cuadrados en dos etapas

Una vez se determina si es posible recuperar los coeficientes de la forma estructural a partir de la forma reducida, será necesario estimarlos. Ya hemos mencionado que si se trata de una ecuación que está sub-identificada no existe forma de encontrar los coeficientes estructurales y por lo tanto la solución es replantear el modelo. Mientras que si se trata de una ecuación perfectamente identificada podemos estimar los coeficientes de forma reducida, y después encontrar, a partir de ellos, los coeficientes estructurales por Mínimos Cuadrados Indirectos, sin embargo esto es demasiado tedioso. Es por esta razón que para ecuaciones perfectamente identificadas y sobre-identificadas emplearemos el método de mínimos cuadrados en dos etapas (MC2E), con el cual podemos eliminar el problema de simultaneidad.

Intuitivamente, la idea detrás de los MC2E es remover la parte aleatoria de la parte determinística de las variables endógenas que actúan como variables explicativas. En otras palabras, este método implicará en un primer paso separar la parte determinística de las variables endógenas explicativas de la parte aleatoria, esto se realiza a partir de la forma reducida del sistema. Y en un segundo paso, se estima la ecuación estructural remplazando la variable endógena explicativa por la parte deterministica de ella. Se elimina así la relación entre una variable aleatoria explicativa y el término de error de ese

modelo. Veamos cómo funciona este método por medio del siguiente ejemplo:

$$Q_t^d = \alpha_1 + \alpha_2 p_t + \alpha_3 Y_t + \varepsilon_t^d \tag{8.14}$$

$$Q_t^S = \beta_1 + \beta_2 p_t + \beta_3 pins_t + \upsilon_t^s \tag{8.15}$$

Mediante la condición de orden podemos darnos cuenta que las ecuaciones 8.14 y 8.15 se encuentran probablemente perfectamente identificadas y mediante la forma reducida obtenemos la siguiente ecuación para p_t:

$$\hat{p}_t = \hat{\pi}_{11} + \hat{\pi}_{12} Y_t + \hat{\pi}_{13} p_{ins} + \mu_t$$

En este caso, los MC2E equivalen a efectuar los siguientes pasos:

1. Encuentre la forma reducida del modelo
2. Encuentre $\hat{p}_t = \hat{\pi}_{11} + \hat{\pi}_{12} Y_t + \hat{\pi}_{13} p_{ins}$
3. Estime la ecuación estructural de la siguiente forma $Q_t^d = \alpha_1 + \alpha_2 \hat{p}_t + \alpha_3 Y_t + \varepsilon_t^d$

Los estimadores MC2E tienen las siguientes propiedades:

- Si el R^2 de la primera regresión ($\hat{p}_t = \hat{\pi}_{11} + \hat{\pi}_{12} Y_t + \hat{\pi}_{13} p_{ins}$) no es muy alto, entonces los MC2E no tienen mucho sentido.
- Son asintóticamente consistentes y eficientes
- No son insesgados en muestras pequeñas.

8.6. Prueba de simultaneidad de Hausman

Recuerden que el problema es causado por la simultaneidad de una variable endógena del modelo (empleada como variable explicativa) y el término de error de esa ecuación. Para entender cómo funciona este test, empleemos el siguiente modelo de oferta y demanda

$$Q_t^s = \alpha_1 + \alpha_2 p_t + \varepsilon_t^s \tag{8.16}$$

$$Q_t^d = \beta_1 + \beta_2 p_t + \beta_3 y_t + \beta_4 w_t + \upsilon_t^d$$

$$Q_t^d = Q_t^S = Q_t \tag{8.17}$$

donde y_t y w_t representan el ingreso y la riqueza respectivamente. Entonces la forma reducida será:

$$p_t = \pi_{11} + \pi_{12} Y_t + \pi_{13} w_t + \mu_t$$

Tras estimar por MCO el anterior modelo, podemos obtener:

$$\hat{p}_t = \hat{\pi}_{11} + \hat{\pi}_{12} y_t + \hat{\pi}_{13} w_t$$

y por lo tanto:

$$p_t = \hat{p}_t + \hat{\upsilon}_t \tag{8.18}$$

Noten que con esto hemos separado la parte aleatoria $\hat{v}_t$ de la determinística $\hat{p}_t$. Por lo tanto, tendremos que reemplazando 8.18 en 8.16 obtenemos:

$$Q_t^s = \alpha_1 + \alpha_2 \hat{p}_t + \alpha_2 \hat{v}_t + \varepsilon_t^s$$

Así, bajo la H_0 de no simultaneidad[4] tendremos que los errores no están correlacionados.

Entonces el test de simultaneidad comprenderá los siguientes pasos:

1. Encuentre $\hat{p}_t$, $\hat{v}_t$:
2. Estime el modelo $Q_t^s = \alpha_1 + \alpha_2 \hat{p}_t + \delta \hat{v}_t + \varepsilon_t^s$
3. Compruebe por medio del $t - cal$ la $H_0 : \delta = 0$

Por lo tanto, si rechazamos H_0 el modelo presenta problemas de simultaneidad y es buena idea emplear el método MC2E para estimar 8.16. Es importante aclarar que si existe mas de un regresor endógeno involucrado, se debe emplear una prueba de significancia conjunta.

8.7. Relación entre la inflación y el nivel de apertura

Para este ejercicio emplearemos el modelo teórico desarrollado por Romer (1993) para explicar las diferencias entre los niveles de inflación de diferentes países alrededor del mundo. En resumen, Romer llega a la conclusión que los países más "abiertos" al comercio internacional deben tener menor inflación. En especial, Romer plantea el siguiente sistema de ecuaciones simultáneas.[5]

$$Inf_i = \beta_0 + \beta_1 apert_i + \beta_2 \ln(ingrpc_i) + \varepsilon_i \tag{8.19}$$

$$apert_i = \alpha_0 + \alpha_1 Inf_i + \alpha_2 \ln(ingrpc_i) + \alpha_3 \ln(exterr_i) + \mu_i \tag{8.20}$$

donde Inf_i corresponde a la inflación promedio anual para el período 1973-1992 para el país i; $apert_i$ denota el grado de apertura del país i medido como el promedio anual, para el período 1973-1992, de las importaciones como porcentaje del PIB; $ingrpc_i$ expresa el ingreso per cápita para el país i, medido como el promedio anual, para el período 1973-1992, del ingreso per. cápita en dólares constantes de 1980; y $exterr_i$ es la extensión territorial del país i, medido en millas cuadradas. Finalmente, ε_i y μ_i corresponden a términos aleatorios de error que no están relacionados entre si.

En el archivo eecsimul.xls usted cuenta con una base de datos para 114 diferentes países. Estime, de ser posible, el sistema de ecuaciones dado por las ecuaciones 8.19 y 8.20. Note que existen dos variables endógenas (Inf_i y $apert_i$) y dos exógenas ($ingrpc_i$ y $exterr_i$). El lector puede darse cuenta fácilmente que la ecuación 8.19 se encuentra perfectamente identificada, mientras que la 8.20 está sub-identificada.

Como lo discutimos en este capítulo, la ecuación 8.19 no se debe estimar directamente por el método de MCO, pues hay una gran probabilidad que los estimadores MCO sean sesgados gracias a la posible

[4] Es decir la variable explicatoria no está correlacionada con el error.

[5] El modelo de la forma estructural se puede rescribir de la siguiente forma reducida:

$$Inf_i = \pi_{11} + \pi_{12} \ln(ingrpc_i) + \pi_{13} \ln(exterr_i) + \mu_{1i}$$

$$apert_i = \pi_{21} + \pi_{22} \ln(ingrpc_i) + \pi_{23} \ln(exterr_i) + \mu_{2i}$$

correlación del término de error y las variables endógenas que aparecen como variables explicativas. Así, no podemos emplear el método de Mínimos Cuadrados Ordinarios para estimar los coeficientes de esta ecuación estructurales, pero podemos emplear el método de MC2E. Por otro lado, la ecuación 8.20 no es posible estimarla por estar sub-identificada.

A continuación, encontrará los pasos a seguir para estimar la ecuación 8.19 por MC2E. Además se discutirán brevemente los pasos necesarios para efectuar el test de simultaneidad de Hausman.

Antes de iniciar, cargue los datos que se encuentran en el archivo eecsimul.xls en EasyReg y cree las variables: $\ln(ingrpc_i)$ y $\ln(exterr_i)$.

8.7.1. Estimación por el método de mínimos cuadrados en dos etapas.

Los MC2E implican dos pasos. Si queremos estimar la ecuación 8.19, primero se debe estimar $apert_i$ a partir de la ecuación de la forma reducida:

$$apert_i = \pi_{21} + \pi_{22} \ln(ingrpc_i) + \pi_{23} \ln(exterr_i) + \mu_{2i}$$

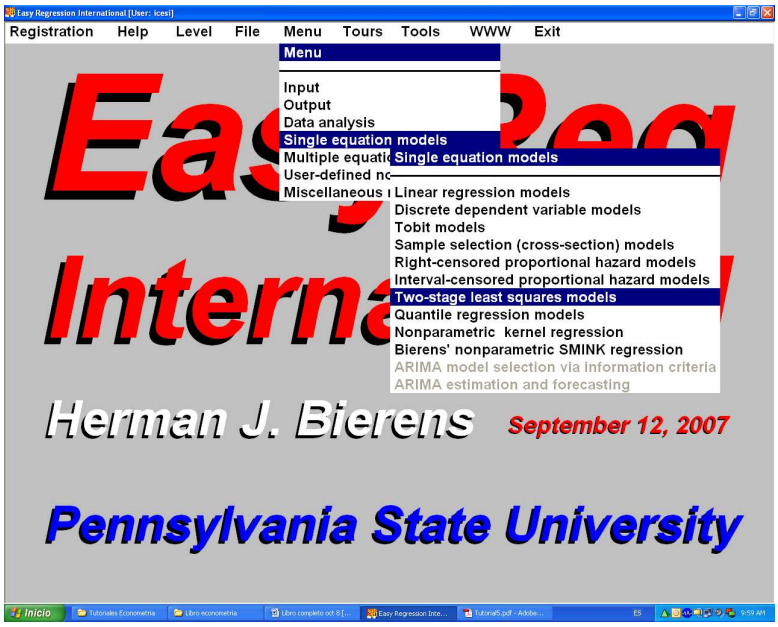
El segundo paso es estimar:

$$Inf_i = \beta_0 + \beta_1 \widehat{apert_i} + \beta_2 \ln(ingrpc_i) + \varepsilon_i$$

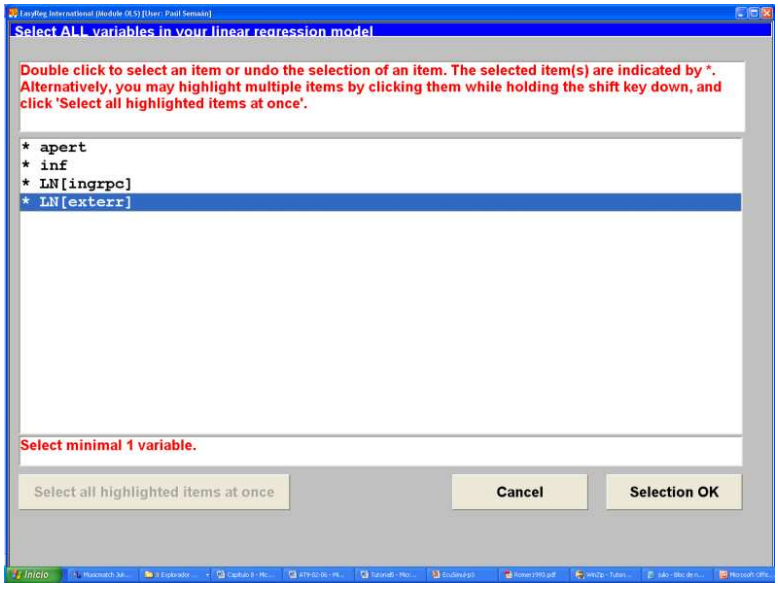
Estos dos pasos son efectuados simultáneamente por EasyReg. A continuación, veremos cómo efectuar una estimación por el método de mínimos cuadrados en dos etapas (también conocido como el método de variables instrumento). Pero antes de continuar distingamos 4 grupos de variables relacionadas con la ecuación 8.19:

1. Variable dependiente (endógena): Inf_i.
2. Variables endógenas explicativas: $apert_i$ (en este caso solo hay una variable pero pueden existir varias).
3. Variables exógenas explicativas: $ln(ingrpc_i)$
4. Variables exógenas del sistema que no se incluyen en esta ecuación: $ln(exterr_i)$

Ahora, para estimar el model por el método MC2E en EasyReg haga clic en "Menu/Single equation models/Two-stage least squares models"



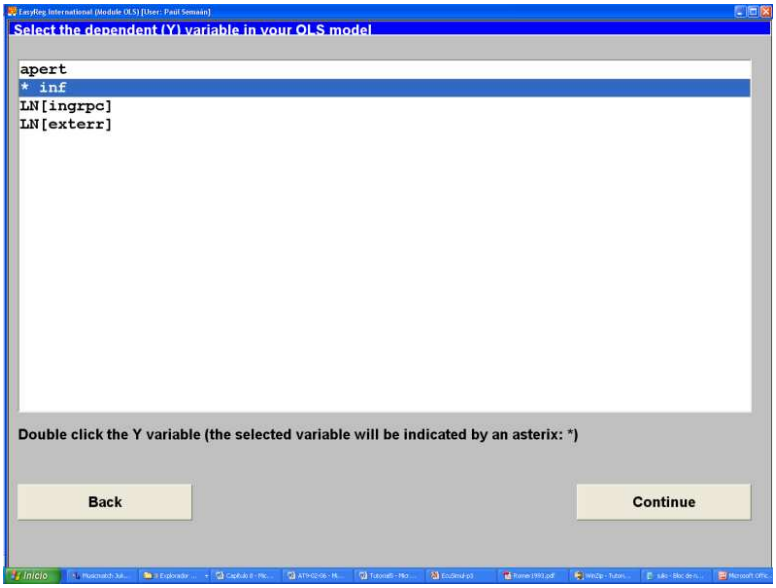
Observará una ventana muy familiar para usted. En esta ventana escoja, haciendo doble clic sobre ellas, todas las variables que están incluidas en el modelo 8.19 y en aquellas variables exógenas que estén en el sistema de ecuaciones simultáneas y que no estén en la ecuación a estimar, es decir Inf_i, $apert_i$, $\ln(ingrpc_i)$ y $\ln(exterr_i)$. Todas las variables descritas en los 4 grupos arriba mencionados.



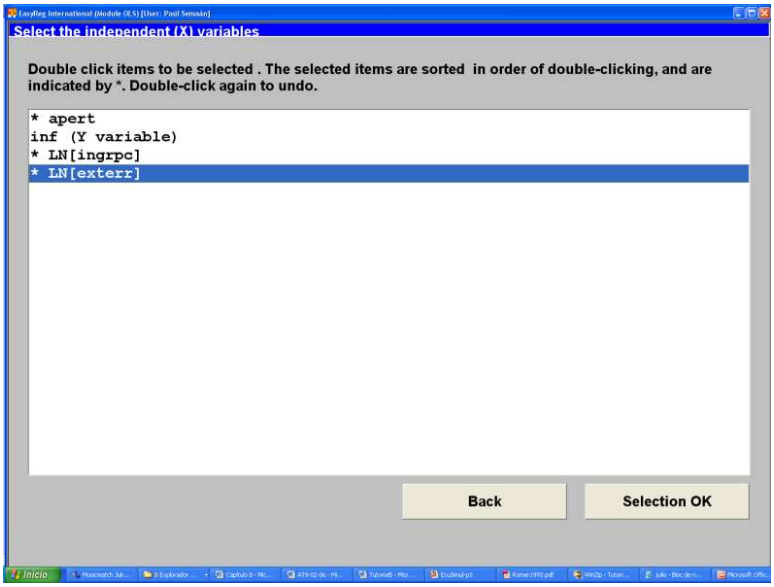
Haga clic en el botón "Selection OK". A continuación haga clic en el botón "No" y posteriormente en el botón "Continue". Observará la siguiente ventana:

Recuerde que se desea estimar la siguiente ecuación:

$$Inf_i = \beta_0 + \beta_1 apert_i + \beta_2 \ln(ingrpc_i) + \varepsilon_i$$



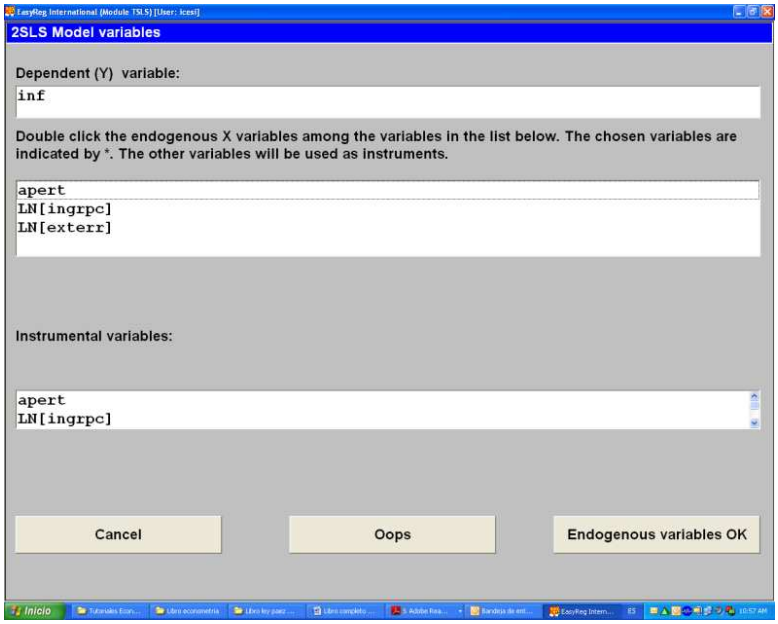
En esta ventana debemos seleccionar la variable dependiente. Escoja la variable "Inf_i*" y haga clic en el botón "Continue" dos veces. Observará la siguiente ventana:



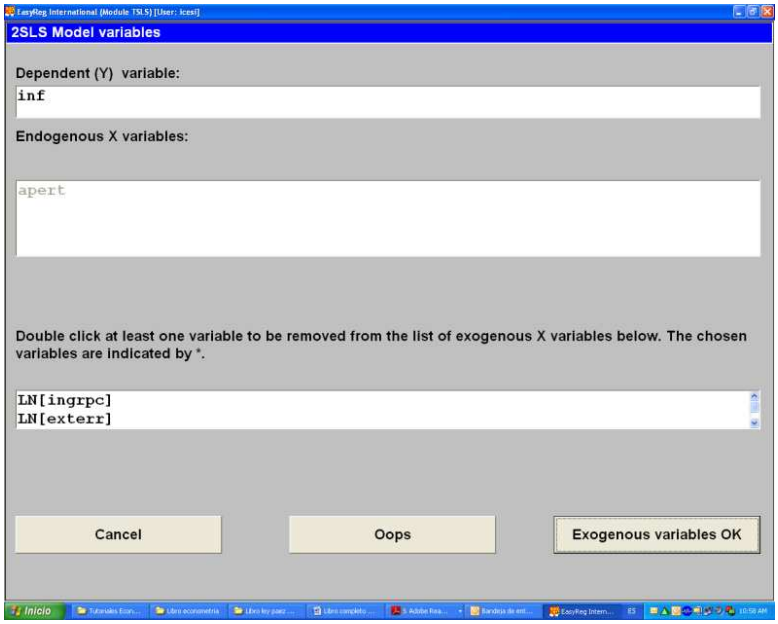
Esta ventana permite seleccionar las variables explicativas para el modelo y las variables que se emplearán como instrumento para la estimación.

En la primera ventana, "Dependent (Y) variable", encontrará la variable dependiente de la ecuación estimada previamente seleccionada. En la segunda ventana, usted debe escoger las variables endógenas incluidas como variables explicativas;[6] en este caso escoja "$apert_i$" y haga clic en el botón "Endogenous variables OK". Observará la siguiente ventana.

[6]Esto permite a Easyreg saber cuál es la variable o variables que deben ser reemplazadas por su valor estimado a partir de la forma reducida del sistema



En la tercera ventana, panel inferior, se deben seleccionar las variables exógenas excluidas de la ecuación 8.19.[7] Para esto, tenemos que hacer clic en la variable $\ln(exterr_i)$ y posteriormente haga clic en el botón "Exogenous variables OK". Observará la siguiente ventana:



Haga clic en el botón "Continue" dos veces. Observará los resultados reportados en el recuadro 8.2

[7]Esto permite a Easyreg determinar cuáles son las variables exógenas que, estarán en la ecuación estructural y aquellas exógenas que si bien no estarán en la forma estructural, sí estarán en la estimación de la forma reducida.

Recuadro 8.2 Estimación de la ecuación 8.19 por MC2E.

Two-stage least squares:
Dependent variable:
Y = inf

Characteristics:
inf
First observation = 1
Last observation = 114
Number of usable observations: 114
Minimum value: 3.6000000E+000
Maximum value: 2.0670000E+002
Sample mean: 1.7264035E+001

X variables, including instrumental variables:
X(1) = apert
X(2) = LN[ingrpc]
X(3) = LN[exterr]
X(4) = 1

Endogenous X variable:
Y*=apert
Exogenous X variables:
X*(1)=LN[ingrpc]
X*(2)=1

2SLS estimation results for Y = inf

Variables	2SLS estimate	t-value [p-value]
apert	-0.337487	-2.342 [0.01920]
LN[ingrpc]	0.375825	0.187 [0.85205]
1	26.899337	1.747 [0.08071]

The p-values are two-sided and based on the normal approximation

Standard error of the residuals = 23.835811E+000
Residual sum of squares (RSS) = 63.064195E+003
Total sum of squares (TSS) = 65.073422E+003
R-square = 0.030876
Adjusted R-square = 0.013415
Effective sample size (n) = 114

If the model is correctly specified, in the sense that the conditional
expectation of the model error u relative to the instrumental variables
equals zero, then the 2SLS parameter estimators b(1),...,b(3),
minus their true values, times the square root of the sample size n (=114),

```
are (asymptotically) jointly normally distributed with zero mean vector
and variance matrix:
  2.36788370705938    -6.92963711058899    -34.7259446158907
 -6.92963711058899     462.902706258368    -3288.33466379246
 -34.7259446158907    -3288.33466379246     27040.4382677868
provided that the conditional variance of the model error u is constant
(u is homoskedastic)
```

Noten que en la salida de Easyreg se puede chequear si se han cometido errores o no, mirando "X variables, including instrumental variables", "Endogenous X variable" y "Exogenous X variables". Los resultados se resumen en el cuadro 8.1.[8]

[8]Es importante tener en cuenta que Romer (1993) obtiene resultados distintos para esta estimación debido a que este autor emplea variables dummy en el modelo.

Cuadro 8.1: Estimación del modelo 8.19

	Variable dependiente inf_t Estadísticos t entre paréntesis
	Ecuación 8.19 1973-1992 MC2O
Constante	26.899 (-1.747)*
$apert_i$	-0.337 (-2.34)**
exp_i	0.03048 (5.21)***
$\ln(ingrpc_i)$	0.375 (0.187)
R^2	0.030
R2 Ajustado	0.013
No. de obs.	114

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

MC2O: Mínimos cuadrados en dos etapas

Funte: EasyReg

8.7.2. Prueba de simultaneidad de Hausman

Recuerde que los estimadores MCO serán sesgados e inconsistentes en presencia de una correlación entre uno de los regresores endógenos al sistema y el término de error de la ecuación estructural (esto se conoce como el problema de simultaneidad) a estimar. Pero en caso de que la simultaneidad no se presente, los Estimadores MCO serán insesgados, consistentes y eficientes.

Así, en la práctica es importante comprobar si existe o no simultaneidad entre el regresor endógeno al sistema y el término de error de la ecuación estructural a estimar. En las siguientes páginas usted aprenderá a realizar el test de Hausman de simultaneidad, cuya hipótesis nula es la presencia de no simultaneidad entre los regresores endógenos y la hipótesis alterna es la presencia de simultaneidad.

En primer lugar, estime la forma reducida de $apert_i$ por el método MCO

$$apert_i = \pi_{21} + \pi_{22} \ln(ingrpc_i) + \pi_{23} \ln(exterr_i) + \mu_{2i} \tag{8.21}$$

Obtendrá los siguientes resultados:

Recuadro 8.3 Estimación de la ecuación 8.21

Dependent variable:
Y = apert

Characteristics:
apert
First observation = 1
Last observation = 114
Number of usable observations: 114
Minimum value: 7.4000000E+000
Maximum value: 1.6380000E+002
Sample mean: 3.7078947E+001

X variables:
X(1) = LN[ingrpc]
X(2) = LN[exterr]
X(3) = 1

Model:
Y = b(1)X(1) + b(2)X(2) + b(3)X(3) + U,
where U is the error term, satisfying
E[U|X(1),X(2),X(3)] = 0.

OLS estimation results

Parameters	Estimate	t-value (S.E.) [p-value]	H.C. t-value (H.C. S.E.) [H.C. p-value]
b(1)	0.5464810	0.366 (1.49324) [0.71439]	0.381 (1.43611) [0.70355]
b(2)	-7.5671027	-9.294 (0.81422) [0.00000]	-6.627 (1.14180) [0.00000]
b(3)	117.0845107	7.388 (15.84830) [0.00000]	6.416 (18.24807) [0.00000]

Notes:
1: S.E. = Standard error
2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
3: The two-sided p-values are based on the normal approximation.

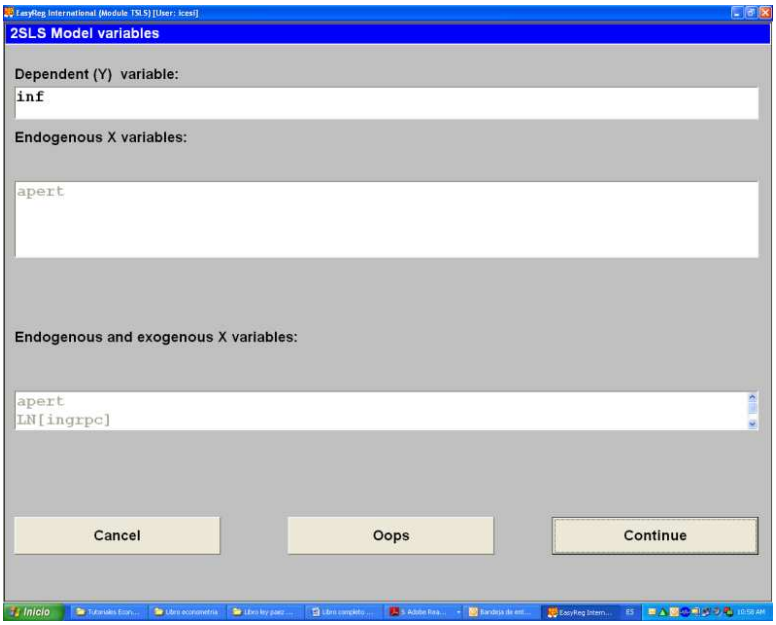
Effective sample size (n): 114
Variance of the residuals: 316.68285702
Standard error of the residuals (SER): 17.79558532
Residual sum of squares (RSS): 35151.79712966

(Also called SSR = Sum of Squared Residuals)
Total sum of squares (TSS): 63757.98947368
R-square: 0.4487
Adjusted R-square: 0.4387

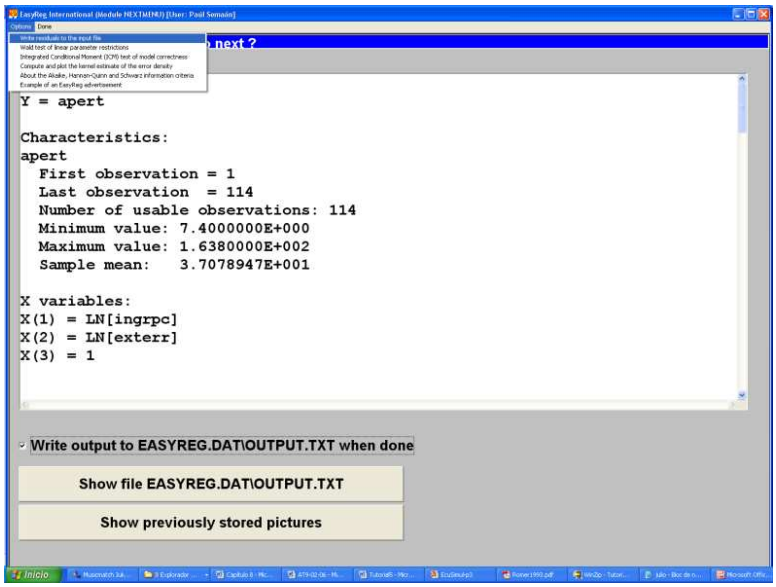
Overall F test: F(2,111) = 45.17
p-value = 0.00000

Significance levels:	10 %	5 %
Critical values:	2.35	3.08
Conclusions:	reject	reject

Al finalizar la estimación en la parte superior elija "Options" y luego haga clic en "Write residuals to the input file"



Posteriormente debemos crear la variable $\widehat{apert_i}$, recuerde que esto es equivalente a $apert_i - \hat{\mu}_{2i}$. Por lo tanto, creemos esta variable mediante la opción "Transform variables" del menú principal. Observará la siguiente ventana:



Luego estimemos el siguiente modelo

$$Inf_i = \beta_0 + \beta_1 \widehat{apert}_i + \beta_2 \ln(\widehat{ingrpc}_i) + \delta \hat{\mu}_i + \varepsilon_i \tag{8.22}$$

Donde $\hat{\mu}_i$ son los errores que creamos cuando estimamos la función $apert_i$ en su forma reducida. Los resultados se presentan en el recuadro 8.4, en la cual es claro que no podemos rechazar la $H_0 : \delta = 0$ y por lo tanto no podemos rechazar la H_0 de no simultaneidad. En otras palabras el Test de Hausman no revela problemas de simultaneidad en el modelo estimado, de tal manera que la estimación de la ecuación 8.19 se debe estimar por MCO y no por MC2E.

Recuadro 8.4 Estimación de la ecuación 8.22

Dependent variable:
Y = inf

Characteristics:
inf
First observation = 1
Last observation = 114
Number of usable observations: 114
Minimum value: 3.6000000E+000
Maximum value: 2.0670000E+002
Sample mean: 1.7264035E+001

X variables:
X(1) = apert-OLS Residual of apert
X(2) = LN[ingrpc]
X(3) = OLS Residual of apert
X(4) = 1

Model:
Y = b(1)X(1) +.....+ b(4)X(4) + U,
where U is the error term, satisfying
E[U|X(1),...,X(4)] = 0.
OLS estimation results

Parameters	Estimate	t-value (S.E.) [p-value]	H.C. t-value (H.C. S.E.) [H.C. p-value]
b(1)	-0.33749	-2.363 (0.14285) [0.01815]	-2.279 (0.14806) [0.02265]
b(2)	0.37582	0.188 (1.99731) [0.85075]	0.281 (1.33662) [0.77858]
b(3)	-0.11981	-0.951 (0.12601) [0.34171]	-2.221 (0.05396) [0.02638]
b(4)	26.89934	1.762 (15.26537) [0.07805]	2.603 (10.33323) [0.00924]

Notes:
1: S.E. = Standard error
2: H.C. = Heteroskedasticity Consistent. These t-values and standard errors are based on White's heteroskedasticity consistent variance matrix.
3: The two-sided p-values are based on the normal approximation.

Effective sample size (n): 114
Variance of the residuals: 558.169197
Standard error of the residuals (SER): 23.625605
Residual sum of squares (RSS): 61398.611647
(Also called SSR = Sum of Squared Residuals)
Total sum of squares (TSS): 65073.422544
R-square: 0.0565
Adjusted R-square: 0.0307

Overall F test: F(3,110) = 2.19
p-value = 0.09275

Significance levels:	10 %	5 %
Critical values:	2.13	2.69
Conclusions:	reject	accept

8.8. Ejercicios

Un centro de estudios económicos ha considerado el siguiente modelo de ecuaciones simultáneas para analizar la economía de una pequeña república caribeña:

$$C_t = \alpha_1 + \alpha_2 Y_t + \alpha_3 C_{t-1} + \alpha_4 T_t + \varepsilon_{1t} \tag{8.23}$$

$$I_t = \beta_1 + \beta_2 Y_{t-1} + \varepsilon_{2t} \tag{8.24}$$

$$T_t = \gamma_1 + \gamma_2 Y_t + \varepsilon_{3t} \tag{8.25}$$

$$M_t = \delta_1 + \delta_2 Y_t + \delta_3 Y_{t-1} + \delta_4 M_{t-1} + \varepsilon_{4t} \tag{8.26}$$

$$Y_t \equiv C_t + I_t + G_t + X_t - M_t \tag{8.27}$$

Donde, C_t, I_t, T_t, M_t, X_t, G_t y Y_t representan el Consumo privado nacional, la Inversión privada nacional, la recaudación directa, las Importaciones, las Exportaciones, el Gasto público y el PIB respectivamente.

Para llevar a cabo el análisis, el centro de estudios económicos ha recogido la información para el periodo 1970 - 1997 que se encuentra en el archivo "ecsimul.xls". Todas las variables están medidas en millones de moneda local (constantes de 1986). Empleando esta información responda las siguientes preguntas.

1. Identifique empleando la condición de orden, cada una de las ecuaciones del sistema (especifique cuáles son las variables endógenas y exógenas del sistema).
2. A partir de la información recogida por el centro de estudios, estime, para el periodo 1970 - 1997, el sistema empleando el método adecuado.
3. Interprete los coeficientes estimados.
4. Determine el efecto de un aumento del gasto público sobre el consumo, la inversión y el PIB.

8.9. Apéndice

Apéndice 8.1 Consecuencia de la correlación entre variables explicativas y el término de error

En esta demostración mostramos cómo en el caso de la estimación de una ecuación de la forma estructural en la que esté presente una variable endógena al modelo como variable explicativa, se presentará una correlación entre el termino de error y las variables explicativas. Es decir, en el caso de una ecuación de la forma estructural con una variable endógena como explicativa, no es plausible seguir suponiedo que las $X's$ son no estocásticas. Esto implica que la matriz de variables explicativas no es estadísticamente independiente del término de error. En otras palabras, no son ortogonales. Formalmente,

$$E\left[\mathbf{X}\varepsilon\right] \neq 0$$

Veamos cómo afecta este hecho a la propiedad de insesgades de los MCO. En este caso,

$$E\left[\hat{\beta}\right] = E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T y}\right]$$

$$E\left[\hat{\beta}\right] = E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\left(\mathbf{X}\beta + \varepsilon\right)\right]$$

Manipulando algebraicamente esta expresión obtenemos:

$$E\left[\hat{\beta}\right] = E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta + \left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\varepsilon\right]$$

$$E\left[\hat{\beta}\right] = E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T X}\beta\right] + E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\varepsilon\right]$$

$$E\left[\hat{\beta}\right] = \beta + E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\varepsilon\right]$$

Noten que en el caso de la estimación de ecuaciones estructurales tendremos un sesgo igual a:

$$E\left[\left(\mathbf{X^T X}\right)^{-1}\mathbf{X^T}\varepsilon\right]$$

Este sesgo no desaparecerá al emplear muestras grandes, pues para que dicho sesgo desaparezca se necesita que $E\left[\mathbf{X}\varepsilon\right] = 0$. Y esto no ocurre en este caso, pues como lo demostramos $E\left[\mathbf{X}\varepsilon\right] \neq 0$.

Parte IV

Respuestas

CAPÍTULO 9

Respuestas

Respuestas capítulo 1

1. El modelo 1.24 estimado se presenta acontinuación:

Cuadro 9.1: Estimación del modelo 1.24

	Variable dependiente I_t Estadísticos t entre paréntesis
	Ecuación 1.24 1994-2003 MCO
Constante	93368.73
	(1.66)
CE_t	80.329
	(3.34)***
CD_t	49.363
	(0.25)
$LDiest_t$	-47.919
	(-0.91)
LEl_t	-145.954
	(-2.36)*
V_t	0.017
	(0.55)
R^2	0.99024
R^2 ajustado	0.97804
F-global	81.17***
No. de obs.	10

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos cuadrados ordinarios

2. La interpretación de los coeficientes estimados es la siguiente:

   $\hat{\alpha}_1$ = 93368.73 Los ingresos que no dependen ni del consume de electricidad, ni del consumo de diésel, ni del número de locomotoras diésel, ni del número de locomotoras eléctricas, ni del número de viajeros corresponden a 93368 millones de dólares.
   $\hat{\alpha}_2$ = 80.32 Un aumento de un millón de Kilovatios/hora en el consumo de electricidad aumentará los ingresos en 80.32 millones de dólares.
   $\hat{\alpha}_3$ = 49.36325 Un aumento de un millón de galones en el consumo de diésel aumentará los ingresos en 49.36 millones de dólares.
   $\hat{\alpha}_4$ = -47.91913 Una locomotora diésel más en el país disminuirá los ingresos en 47.92 millones de dólares.
   $\hat{\alpha}_5$ = -145.954 Una locomotora eléctrica más en el país disminuirá los ingresos en 145,954 millones de dólares.
   $\hat{\alpha}_6$ = 0.017444 Mil viajeros más aumentarán los ingresos en 17,444 dólares.

Respuestas capítulo 2

1. Para determinar cuál es el mejor modelo de los ocho a considerar, podemos emplear el R^2 ajustado de cada modelo. De acuerdo con estos resultados, el mejor modelo corresponde al modelo 2.31, pues es el que posee el R^2 ajustado mayor entre todos los modelos.

Cuadro 9.2: Resultados de las estimaciones

	Variable dependiente I_t Estadísticos t entre paréntesis							
	Ecuación 1.24 1994-2003 MCO	Ecuación 2.26 1994-2003 MCO	Ecuación 2.27 1994-2003 MCO	Ecuación 2.28 1994-2003 MCO	Ecuación 2.29 1994-2003 MCO	Ecuación 2.30 1994-2003 MCO	Ecuación 2.31 1994-2003 MCO	Ecuación 2.32 1994-2003 MCO
Constante	93368.73	-31144.44	308808.10	71774.82	74977.75	260199.00	85165.48	-21168.96
	(1.66)	(0.79)	(8.39)***	(10.03)***	(1.76)	(16.12)***	(2.08)*	(-0.49)
CE_t	80.329	121.106	–	–	100.249	–	83.504	105.727
	(3.34)***	(7.12)***			(6.96)***		(4.54)***	(3.83)***
CD_t	49.363	-375.413	–	–	–	327.112	–	-315.721
	(0.25)	(-4.07)***				(-4.07)***		(-2.50)**
$LDies_t$	-47.919	–	-120.300	–	–	-269.094	-38.091	–
	(-0.91)		(-1.99)**			(-3.38)**	(-1.23)	
LEl_t	-145.954	–	-175.021	–	-207.963	–	-140.370	–
	(-2.36) *		(-1.44)		(-6.13)***		(-2.71)**	
Vt	0.017	–	–	0.201	–	–	0.016	0.031
	(0.55)			(8.06)***			(0.57)	(0.72)
R^2 0.99024	0.97069	0.92136	0.89029	0.98449	0.90842	0.99009	0.97304	
R^2 ajustado	0.9780	0.9623	0.8989	0.8766	0.9801	0.8823	0.9822	0.9596
F-global	81.17***	115.92***	41.01***	64.92***	222.11***	34.72***	124.88***	72.19***
No. de Obs.	10	10	10	10	10	10	10	10

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos cuadrados ordinarios

2. *a*) La tabla Anova del modelo seleccionado es la siguiente:

Cuadro 9.3: Tabla Anova modelo 2.31

Fuente de la Variación	**SS**	**Grados de libertad**	**MS**	**F-global**
Regresión	2,510,720,956.3	4	627680239	124.8806057
Error	25131213.8	5	5026243	
Total	2535852170.1	9		

b) Para probar individualmente la hipótesis nula de que $\alpha_i = 0$ versus la hipótesis alterna $\alpha_i \neq 0$ para $i = 1, 2..,5$ podemos remitirnos a los estadísticos t. Como se puede observar, individualmente $\hat{\alpha}_3$ y $\hat{\alpha}_5$ no son significativos. Por otra parte, los coeficientes asociados al consumo de energía y el número de locomotoras eléctricas son significativos a un nivel de significancia del 1 % y 5 %, respectivamente.

c) Para comprobar la hipótesis nula de que $\alpha_2 = \alpha_3 = ... = \alpha_5 = 0$ versus la hipótesis alterna no H_0, se puede emplear el estadístico F-global que en este caso corresponde a 124.8806

el cual es mayor al valor crítico de la distribución F (11.39) con 4 y 5 grados de libertad en el numerador y denominador, respectivamente. Por tanto se puede rechazar la hipótesis nula de que todos los coeficientes asociados a pendientes son conjuntamente iguales a cero. Luego los coeficientes son conjuntamente significativos.

d) Para poder efectuar esta comparación, debemos emplear los coeficientes estandarizados que se calculan de la siguiente forma: $\hat{\alpha}_i^E = \hat{\alpha}_i \frac{s_{Xi}}{s_y}$ para $i = 2, 3, ..., 5$. En el cuadro se presentan los resultados de estandarizar los coeficientes. Noten que el coeficiente estandarizado más grande es el asociado con el consumo de energía. Así, esta es la variable que afecta más el ingreso ferroviario.

Cuadro 9.4: Coeficientes estandarizados del modelo 2.31

Variable	**Coeficiente**		**Desviación**
	MCO	**Estandarizado**	**estándar**
I_t	NA	NA	16785.74849
CE_t	-38.091	-0.179	78.96447865
$LDies_t$	-140.370	-0.330	39.43616502
V_t	0.016	0.077	78739.52557

NA: No Aplica
MCO: Mínimos Cuadrados Ordinarios

e) Noten que parece paradójico que a mayor consumo de energía mayores ingresos, pero esto puede estar asociado a que las locomotoras eléctricas son las que generan más ingresos para el sector. Además, parece razonable concluir, dado el bajo coeficiente estandarizado asociado a los viajeros, que los ingresos son generados por transporte de mercancías. Así, será importante para el gobierno tener en cuenta los efectos del transporte de mercancías para este sector. Sería conveniente replantear el modelo, empleando información de carga transportada por este sector para poder llegar a recomendaciones más concretas.

Respuestas capítulo 3

1. Aunque la serie presenta alta volatilidad, es apreciable que los efectos de ésta dentro del periodo estudiado anulan sus propias variaciones, es decir sus efectos contrarios se eliminan, observándose que aparentemente la media de la serie de los rendimientos es igual a cero.

Figura 9.1: Serie de rendimientos de la TRM

Fuente: Superintendencia Bancaria

2. El modelo que captura esta situación está dado por:

$$Rend_t = \alpha_1 + \alpha_2 Martes_t + \alpha_3 Miercoles_t + \alpha_4 Jueves_t + \alpha_5 Viernes_t + \varepsilon_t \tag{9.1}$$

Donde $Martes_t$, $Mircoles_t$, $Jueves_t$ y $Viernes_t$ son variables ficticias definidas de la siguiente manera:

$$Martes_t = \begin{cases} 1,\ t \in \text{ Martes} \\ 0,\ t \in \text{ o.w} \end{cases} \qquad Miercoles_t = \begin{cases} 1,\ t \in \text{ Mircoles} \\ 0,\ t \in \text{ o.w} \end{cases}$$

$$Jueves_t = \begin{cases} 1,\ t \in \text{ Jueves} \\ 0,\ t \in \text{ o.w} \end{cases} \qquad Viernes_t = \begin{cases} 1,\ t \in \text{ Viernes} \\ 0,\ t \in \text{ o.w} \end{cases}$$

Entonces:

$$Rend_t = \begin{cases} (\alpha_1 + \alpha_2)\, Martes_t + \varepsilon_t & \text{t = Martes} \quad \cdots \\ (\alpha_1 + \alpha_3)\, Miercoles_t + \varepsilon_t & \text{t = Miercoles} \cdots \\ (\alpha_1 + \alpha_4)\, Jueves_t + \varepsilon_t & \text{t = Jueves} \quad \cdots \\ (\alpha_1 + \alpha_5)\, Viernes_t + \varepsilon_t & \text{t = Viernes} \quad \cdots \\ (\alpha_1) + \varepsilon_t & \text{t = Lunes} \quad \cdots \end{cases}$$

El término independiente recoge el valor medio del rendimiento observado en el día lunes del periodo de estudio, mientras el coeficiente α_2 es el diferencial existente que corresponde al valor del rendimiento medio del día martes, α_3, es el diferencial que corresponde al valor del rendimiento medio del día miércoles, y así sucesivamente.

3. En el siguiente cuadro se presentan las estimaciones de los modelos

Cuadro 9.5: Resultados de las estimaciones

	Variable dependiente $Rend_t$ Estadísticos t entre paréntesis	
	Ecuación 9.1	Ecuación 9.2
	MCO	MCO
Constante	0.0006	0.0002
	(1.10)	(0.72)
$Martes_t$	-0.0013	-0.0009
	(-1.85)*	(-1.74)*
$Miercoles_t$	-0.0010	-
	(1.50)	-
$Jueves_t$	-0.0004	-
	(0.54)	-
$Viernes_t$	-0.0001	-
	(0.20)	-
R^2	0.01900	0.0095
R^2 ajustado	0.00630	0.0064
F-global	1.50	-
No. de Obs.	315	315

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos cuadrados ordinarios

4. A partir de los resultados anteriores se puede observar que el único parámetro estadísticamente significativo de manera individual es el que acompaña a la variable ficticia del día martes, entonces se procede a evaluar la significancia conjunta de los parámetros que acompañan a las variables que representan al resto de los días de la semana α_3, α_4 y α_5.

   Para llevar a cabo esta prueba de hipótesis conjunta se plantea la siguiente hipótesis nula: $\alpha_3 = \alpha_4 = \alpha_5 = 0$, contra no H_0.

   El estadístico de Wald es igual a 2.99 que se contrasta con un estadístico de prueba que sigue una distribución χ^2 con 3 grados de libertad, al 5 % este valor corresponde a 7.81 y al 10 % corresponde a 6.25, por lo tanto no se puede rechazar la hipótesis nula de no significancia conjunta.

5. Entonces se puede estimar un modelo en el que se omiten las variables que representan los rendimientos del día miércoles, jueves y viernes, como se muestra a continuación:

$$Rend_t = \beta_0 + \beta_1 Martes_t + \upsilon_t \tag{9.2}$$

   Los resultados de la estimación se resumen en el cuadro 9.5.

6. $\hat{\beta}_0 = 0{,}0002$ Corresponde al rendimiento de los días hábiles de la semana promedio, excepto el rendimiento observado en el día martes, no es estadísticamente significativo al 10 %.
   $\hat{\beta}_1 = -0{,}0009$ Corresponde al diferencial del rendimiento de la TCRM cuando el día es martes, es estadísticamente significativo al 10 %.

7. Una aproximación para determinar si es posible sacar ventaja del mercado cambiario es mediante una prueba de hipótesis en el signo y la significancia de los parámetros.

   Para verificar si el signo del parámetro que acompaña a la variable ficticia del día martes es significativamente menor a cero, la hipótesis nula será: $\beta_1 \geq 0$, vs. La hipótesis alterna No H_0. Se compara el t estadístico (-1.74) con los valores críticos z al 10 % (-1.28), al 5 % (-1.64) y al 1 % (-2.32), y por lo tanto se puede rechazar la hipótesis nula con un nivel significancia del 5 %.

   Sobre el signo observado se puede concluir que el parámetro que acompaña al rendimiento del día martes, al ser negativo, es un indicio de la presencia de arbitraje, es decir no existe eficiencia en el mercado cambiario.

   Esta situación hace posible que sea rentable comprar dólares el día martes y venderlos en cualquier otro día de la semana.

Respuestas capítulo 4

1. En el siguiente cuadro se presentan las estimaciones de los modelos

Cuadro 9.6: Resultados de las estimaciones

	Variable dependiente Y_i Estadísticos t entre paréntesis		
	Ecuación 4.9 MCO	Ecuación 4.10 MCO	Ecuación 4.11 MCO
Constante	0.0017	-0.0406	0.0851
	(0.03)	(-0.70)	(8.36)***
$X1_i$	1.39267	2.0588	
	(1.45)	(2.13)**	
$X2_i$	0.08786		0.1138
	(1.85)		(2.48)**
R^2	0.42300	0.25840	0.3214
$R^2 ajustado$	0.3268	0.20140	0.2692
F 4.40**			
No. de Obs.	27	27	27

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos cuadrados ordinarios

Claramente, sí existen síntomas de multicolinealidad pues los t-calculados son relativamente pequeños mientras el F es relativamente alto. Noten que el R^2 no parece ser tan grande. Aun así parece existir problemas de multicolinealidad.

2. **Matriz de correlación de las variables independientes**: Al calcular el determinante de la matriz de correlación $|R|$ por medio de los valores propios se obtiene que el determinante es 0.8588; lo anterior muestra que según esta prueba no parece existir multicolinealidad.

   Medida de Besley, Kuck y Welsch (1980): Dado que los valores propios de la matriz de correlación son 1.3758 y 0.6242; el valor de Kappa equivale a 1.48454623.

$$k(x) = \frac{\sqrt{1.3758}}{\sqrt{0.6242}} = 1.48454623$$

   La anterior medida indica que posiblemente no hay multicolinealidad.

   Matriz de correlaciones de los coeficientes estimados: La matriz de correlaciones de los valores estimados corresponde a:

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$
$\hat{\beta}_1$	1	-0.37576	**-0.9858026**
$\hat{\beta}_2$		1	0.393134
$\hat{\beta}_0$			1

En la matriz se puede observar que el coeficiente estimado para el producto interno bruto presenta una correlación negativa muy alta con el término intercepto; sin embargo, no debe descartarse la información correspondiente a la correlación entre los coeficientes asociados a las variables independientes: Tasa de crecimiento de producto interno bruto y la tasa de inflación.

3. Los resultados se registran en el cuadro 9.6, En las regresiones simples, la significancia de los coeficientes estimados aumentan al nivel de 5 %; lo anterior indica que la regresión original contenía multicolinealidad.

4. A pesar de que en las regresiones simples, la significancia tanto de la tasa de crecimiento del PIB como la tasa de inflación aumenta hasta el 5 %, la supresión de cualquiera de estas variables en la regresión da lugar a estimadores MCO sesgados. Además, la teoría económica sugiere que tanto el PIB como los precios internos deben estar incluidos en la función de importaciones. En conclusión, se debe aceptar la existencia de multicolinealidad, pues esta es imposible de corregir dada la restricción teórica.

Respuestas capítulo 5

1. En el siguiente cuadro se presentan las estimaciones de los modelos

Cuadro 9.7: Resultados de las estimaciones

	Variable dependiente Y_i Estadísticos t entre paréntesis	
	Ecuación 5.11 MCO	Ecuación 9.3 MCP
Constante	-85,184.05	-1,054.0525
	(-2.93)**	NA
X_i	1.1672	1.0901
	(91.52)***	(72.18)***
R^2	0.99809	0.00221
R^2 ajustado	0.99797	-0.06015
F-global	8,375.21***	NA
BP	5.01**	0.0015
No. de Obs.	18	18

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

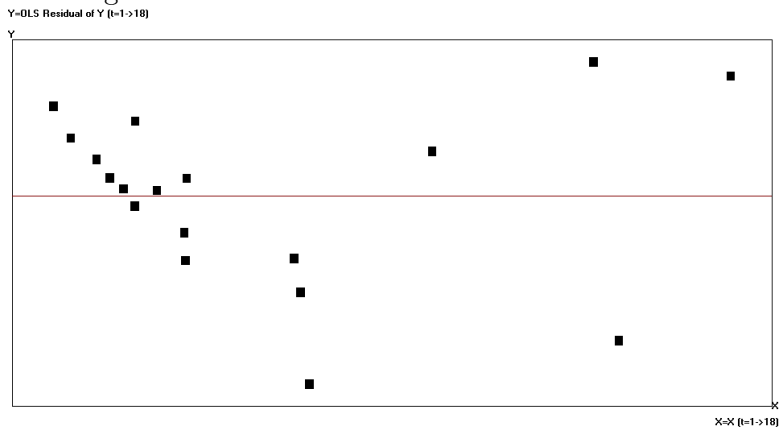
MCO: Mínimos cuadrados ordinarios

MCP: Mínimos cuadrados ponderados

NA: No aplica

2. Inicialmente se grafican los residuos estimados versus la variable X_i

Figura 9.2: Residuos estimados versus renta bruta



Se observa una relación positiva entre la variabilidad de los residuos y la renta bruta. De acuerdo con esto se lleva a cabo la prueba de Goldfeld y Quandt (GQ) planteando la siguiente hipótesis

nula:

$$H_0 : \sigma_i^2 = \sigma^2$$
$$H_A : \sigma_i^2 = \sigma^2 X_i^2$$

El siguiente paso es ordenar las observaciones de acuerdo con la variable que genera el problema de heteroscedasticidad y posteriormente se indican las observaciones que forman las muestras empleadas para calcular el estadístico GQ. En este caso se eliminan 2 observaciones. La primera muestra contiene las observaciones 1 a 8 y la segunda muestra, las observaciones 11 a 18, es decir, se omiten las observaciones 9 y 10. Al estimar las dos regresiones a partir de las dos sub.-muestras, se calcula el estadístico GQ:

$$GQ = \frac{SSE_2}{SSE_1} = \frac{6.8047 \times 10^{10}}{4.2111 \times 10^9} = 16.1572$$

Se compara con $F_{(n-d-2k,n-d-2k)\alpha=0.01} = F_{(12.12)\alpha=0.01} = 4.1552$. Dado que el estadístico GQ es mayor al valor crítico con 12 grados de libertad en el numerador y 12 grados de libertad en el denominador, se rechaza la hipótesis nula a favor de la hipótesis alterna de heteroscedasticidad provocada por la variable X (renta bruta).

EasyReg genera automáticamente un test de Breusch-Pagan asumiendo que todas las variables independientes afectan la varianza del error. En este caso, dado que solo se trata de una variable, se debe obtener los mismos resultados. De acuerdo a esto, para comprobar si la variable afecta la varianza del error, se plantean las siguientes hipótesis:

$$H_0 : \sigma_i^2 = \sigma^2$$
$$H_A : \sigma_i^2 = f\left(\gamma + \delta X_i\right)$$

Con el fin de correr la regresión auxiliar se debe hallar $\hat{\sigma}^2$. En este caso, tenemos que: $SSE = \hat{\varepsilon}^T\hat{\varepsilon} = 1.0755 \times 10^{11}$ y así, $\hat{\sigma}^2 = \frac{\hat{\varepsilon}^T\hat{\varepsilon}}{n} = \frac{1.0755 \times 10^{11}}{18} = 5.9752 \times 10^9$. La regresión auxiliar será entonces:

$$\frac{\hat{\varepsilon}_i^2}{\hat{\sigma}^2} = \gamma + \delta X_i + \mu_i$$

A partir de la regresión auxiliar se obtiene: $SSR = SST - SSE = 30.1770 - 20.1613 = 10.0158$. Ahora se construye el estadístico BP:

$$BP = \frac{SSR}{2} = \frac{10.0158}{2} = 5.0079$$

El cual se compara con $\chi^2_{g(\alpha)} = \chi^2_{1(\alpha)}$. En este caso g representa el número de variables en Z, es decir, una variable. Los valores críticos obtenidos son 6.63, 3.84 y 2.70 para los niveles de significancia del 1 %, 5 % y 10 % respectivamente. De acuerdo con esto, la hipótesis nula se puede rechazar al 5 %, en favor de la presencia de heteroscedasticidad generada por la variable X. Este resultado es igual al generado automáticamente por EasyReg.

El test de White es mucho más general:

$$H_0 : \sigma_i^2 = \sigma^2$$
$$H_A : No\ H_0$$

Para llevar a cabo esta prueba hay que efectuar una regresión auxiliar donde la variable dependiente sean los residuos al cuadrado de la ecuación inicial y las variables independientes sean la variable original y su cuadrado:

$$\hat{\varepsilon}_i^2 = \gamma + \delta_1 X_i + \delta_2 X_i^2 + \upsilon_i$$

El R^2 obtenido a partir de la regresión auxiliar es 0.344573. Ahora se puede construir el estadístico de White: $W_a = nR^2 = 18 \times 0.344573 = 6.2023$ el cual se debe comparar con $\chi^2_{g(\alpha)} = \chi^2_{2(\alpha)}$. Los valores críticos obtenidos son 9.21, 5.99, 4.60 para los niveles de significancia del 1 %, 5 %, 10 % respectivamente. Según a los valores críticos obtenidos, la hipótesis nula se rechaza al nivel de significancia del 5 %. Se concluye entonces que la varianza de los errores no permanece constante a lo largo de la muestra.

3. De acuerdo con el análisis gráfico y las pruebas realizadas se determina que la ecuación 5.11 presenta problemas de heteroscedasticidad. Dado que se puede emplear el método de Mínimos Cuadrados Ponderados (MCP). Conociendo el comportamiento de la varianza de los errores se puede transformar el modelo 5.11 de la siguiente forma:

$$\frac{Y_i}{X_i} = \beta_1 \frac{1}{X_i} + \beta_2 + \frac{\varepsilon_i}{X_i}$$

Reordenando se obtiene:

$$\frac{Y_i}{X_i} = \beta_2 + \beta_1 \frac{1}{X_i} + \frac{\varepsilon_1}{X_i} \tag{9.3}$$

En esta ecuación, el problema de heteroscedasticidad queda resuelto ya que:

$$Var\left[\frac{\varepsilon_i}{X_i}\right] = \frac{1}{X_i^2} Var\left[\varepsilon_i\right] = \frac{1}{X_i^2}\sigma^2 X_i^2 = \sigma^2$$

Los resultados de la estimación de la ecuación 9.3 son reportados en el cuadro 9.7. Como se observa, según el estadístico BP arrojado por EasyReg, el modelo es homoscedástico.

4. β_1= -1054.05: El consumo que no depende de la renta bruta es de -1054.05 millones de pesetas. β_2= 1.09: Un aumento de un millón de pesetas en la renta bruta generará un incremento de 1.09 millones de pesetas en el consumo. Es significativo al 1 %.

Respuestas capítulo 6

1. La ecuación a estimar sería:

$$CONS_t = \alpha + \beta RENT_t + \varepsilon_t \tag{9.4}$$

Cuadro 9.8: Resultados de las estimaciones

	Variable dependiente $CONS_t$ Estadísticos t entre paréntesis	
	Ecuación 9.4 1959:1-2002:2 MCO	Ecuación 9.5 1959:1-2002:2 MCG
Constante	20.17671 (12.916)***	26.6229 n.a.
$RENT_t$	0.0129 (29.469)***	0.01097 (5.351)***
R^2	0.8339	0.1427
R^2 ajustado	0.8329	0.1378
DW	0.1774	2.7072
No. de Obs.	175	174

(*) nivel de significancia: 10 %

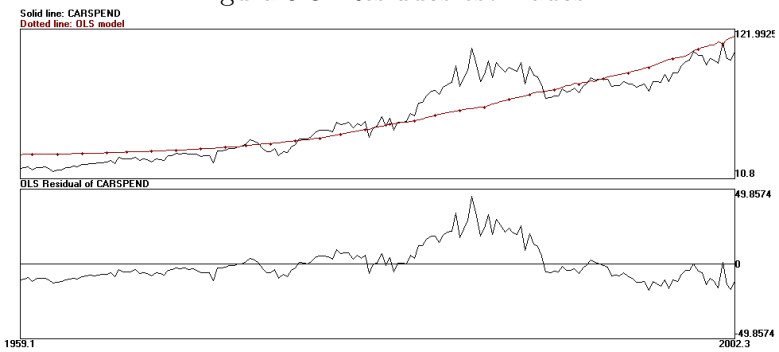
(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos cuadrados ordinarios

MCG: Mínimos cuadrados generalizados

Figura 9.3: Residuos estimados



2. Según lo observado en la figura 9.3, parece existir un problema de autocorrelación pues los errores tienden a seguir un comportamiento. Particularmente, puede ser autocorrelación positiva ya que existen muchos residuos de signo positivo que son seguidos por errores de igual signo. Sin embargo, serán necesarias más pruebas para asumir una conclusión.

3. **Prueba de Rachas**
La hipótesis nula es la existencia de no autocorrelación ($H_0 : \rho = 0$), versus la hipótesis alterna

de autocorrelación ($H_A : \rho \neq 0$). Para la ecuación 9.4, se tiene que k = 15, $N_+ = 65$ y $N_- = 110$. Entonces se hace necesario conocer $E[k] = 82{,}71$ y $Var[k] = 37{,}90$ con estos datos podemos hallar el estadístico $RA = \frac{k-E(k)}{\sqrt{Var(k)}} = -10{,}998$. La regla de decisión es rechazar la hipótesis nula si $|RA| > z_{\alpha/2}$. Si deseamos contrastar a un nivel de significancia del 5 % y el 1 %, tendríamos los valores críticos 1.96 y 2.575 respectivamente, por lo que podemos rechazar la hipótesis nula de y por lo tanto aceptar la existencia de autocorrelación con una confianza del 95 % y 99 %.

Prueba de Durbin Watson

El estadístico de Durbin-Watson corresponde a 0.177350 y los valores críticos son $d_L = 1{,}611$ y $d_u = 1{,}664$ para $n = 150$ y $d_L = 1{,}664$ y $d_u = 1{,}684$ para $n = 200$. Así, para comprobar la $H_0 : \rho = 0$ (no autocorrelación de primer orden) versus la alterna de $H_A : \rho \neq 0$ (existe autocorrelación de primer orden) se debe comparar el DW con d_u y $4 - d_u = 2{,}363$, como 0.177350 es menor que $d_u = 1{,}611$, entonces se puede rechazar la hipótesis nula a favor de la existencia de autocorrelación, lo mismo sucede para el caso de $n = 200$.

Ahora, teniendo en cuenta que 0.177350 es un valor cercano a cero, se considera la H_0 (no existe autocorrelación positiva de primer orden) versus la alterna de $H_A : \rho > 0$. Para rechazar esta hipótesis nula se debe tener que $0 < DW < d_l$, lo cual es cierto para este caso. Por lo tanto, se puede rechazar la hipótesis nula a favor de la hipótesis de la presencia de autocorrelación positiva.

Prueba de Box-Pierce

La hipótesis nula es la no existencia de autocorrelación, versus la hipótesis alterna de la existencia de la misma. De esta manera, los diez primeros rezagos que se muestran en el cuadro 9.9, nos permiten rechazar la hipótesis nula, con lo cual se concluye la existencia de este problema en el modelo.

Cuadro 9.9: Prueba de Box-Pierce

Q(p)	valor	p-valor	valor crítico con el 5 % de significancia	decisión	valor crítico con el 10 % de significancia	decisión
1	143.71	0.00000	3.84	rechazar	2.71	rechazar
2	279.64	0.00000	5.99	rechazar	4.61	rechazar
3	406.49	0.00000	7.81	rechazar	6.25	rechazar
4	525.62	0.00000	9.49	rechazar	7.78	rechazar
5	636.50	0.00000	11.07	rechazar	9.24	rechazar
6	734.77	0.00000	12.59	rechazar	10.64	rechazar
7	821.61	0.00000	14.07	rechazar	12.02	rechazar
8	899.66	0.00000	15.51	rechazar	13.36	rechazar
9	966.70	0.00000	16.92	rechazar	14.68	rechazar
10	1027.79	0.00000	18.31	rechazar	15.99	rechazar

Prueba de Breusch-Godfrey

Para contrastar que los errores de la regresión obedecen a un proceso $AR(1)$ estimamos:

$$\hat{\varepsilon}_t = \beta_0 + \alpha_1 RENT_t + \omega_1 \hat{\varepsilon}_{t-1} + \xi$$

Para un proceso $AR(2)$:

$$\hat{\varepsilon}_t = \beta_0 + \alpha_1 RENT_t + \omega_1 \hat{\varepsilon}_{t-1} + \omega_2 \hat{\varepsilon}_{t-2} + \xi$$

Y así sucesivamente

De cada una de las anteriores estimaciones empleamos el R^2 para calcular el estadístico LM:

$$LM = (n - s) \times R^2$$

Posteriormente comparamos el estadístico LM con el valor crítico de la distribución Chi-cuadrado con s grados de libertad. Rechazaremos la hipótesis nula sí $LM > \chi^2_{s,\alpha}$. Estos resultados se resumen en el cuadro 9.10:

Cuadro 9.10: Resultados de las pruebas de Breush-Godfrey

	Proceso Autorregresivo del Error				
	$AR(1)$	$AR(2)$	$AR(3)$	$AR(4)$	$AR(5)$
R^2	0.8301	0.8539	0.8548	0.8556	0.8552
LM	143.7***	146.0***	144.5***	142.9***	141.1***

(*) Rechaza la H_0 con un nivel de significancia: 10 %

(**) Rechaza la H_0 con un nivel de significancia: 5 %

(***) Rechaza la H_0 con un nivel de significancia: 1 %

4. *a*) De acuerdo con las pruebas realizadas y teniendo en cuenta el análisis gráfico, se llega a la conclusión de que existe un problema de autocorrelación positiva en este modelo. Entonces, podemos utilizar el método de los Mínimos Cuadrados Generalizados. Para ello, rezagamos cada una de las variables un periodo, las multiplicamos por su coeficiente de correlación y se las restamos a las variables no rezagadas, de esta manera se obtiene la ecuación 9.5.

$$CONS^*_t = \alpha^* + \beta RENT^*_t + \varepsilon^*_t \tag{9.5}$$

En donde, $\varepsilon^*_t = \varepsilon_t + \rho\varepsilon_{t-1}$, $CONS^*_t = CONS_t - \rho CONS_{t-1}$, $\alpha^* = \alpha(1-\rho)$ y $RENT^*_t = RENT_t - \rho RENT_{t-1}$.

Primero, se encontró que $\hat{\rho} = 0{,}9236$, el cual resulta después de estimar el modelo:

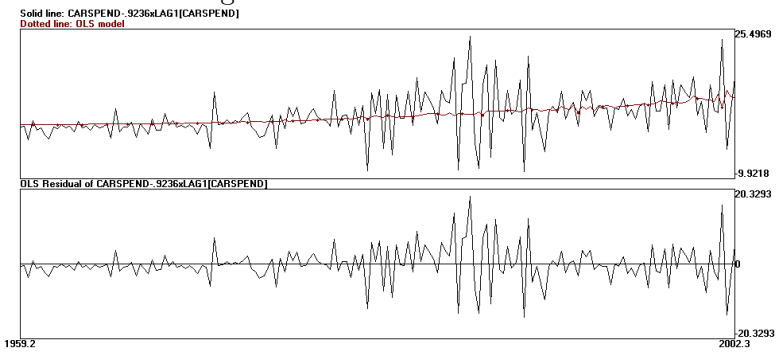
$$CONS_t = \alpha + \beta RENT_t + \rho CONS_{t-1} + \theta RENT_{t-1} + \varepsilon_t$$

Realizados estos pasos, los errores ya no presentarían el problema de autocorrelación en el modelo 9.5, los resultados se reportan en el cuadro 9.9.

b) Nuevamente realizamos las pruebas de Rachas, Durbin-Watson, Box-Pierce y el método gráfico, para comprobar la desaparición del problema.

Análisis gráfico
De la figura 9.4, podemos observar que ahora no podemos detectar un comportamiento en los errores pues unas veces mantienen el signo, como también cambian constantemente, por lo tanto, según este método el problema se ha solucionado. Sin embargo, es necesario realizar otras pruebas que nos permitan sacar mejores conclusiones al respecto.

Figura 9.4: Residuos estimados



Prueba de Rachas

Nuevamente se tiene que la hipótesis nula es la presencia de no autocorrelación ($H_0 : \rho = 0$), versus la hipótesis alterna de autocorrelación ($H_A : \rho \neq 0$). Para la ecuación 9.5, se tiene que k = 89, $N_+ = 71$ y $N_- = 103$. Entonces se hace necesario conocer $E[k] = 85{,}05$ y $Var[k] = 40{,}35$ con estos datos podemos hallar el estadístico $RA = \frac{k-E(k)}{\sqrt{Var(k)}} = 0{,}6206$. La regla de decisión es rechazar la hipótesis nula si $|RA| > z_{\alpha/2}$. Si deseamos contrastar a un nivel de significancia del 5 % y el 1 %, tendríamos los valores críticos 1.96 y 2.575 respectivamente, entonces no podemos rechazar la hipótesis nula de la no existencia de autocorrelación con un nivel de confianza del 95 % y el 99 %, de esta manera se comprueba la corrección del problema.

Prueba de Durbin Watson

En este caso, utilizamos el $DW = 2{,}707165$ obtenido en Easyreg. La hipótesis nula de no autocorrelación que se comprueba por medio del intervalo $d_u < DW < 4 - d_u$ seria igual a: $1{,}637 < DW < 2{,}363$, para $n = 150$ ya que esto no se cumple, entonces se rechaza la hipótesis nula de la inexistencia de autocorrelación, igual caso sucede cuando $n = 200$, se confirma la presencia de autocorrelación negativa.

Prueba de Box-Pierce

En la el cuadro 9.11 se muestran los resultados para los 10 primeros rezagos. De esta manera se rechaza la hipótesis nula de no autocorrelación para estos rezagos.

Cuadro 9.11: Prueba de Box-Pierce

Q(p)	valor	p-valor	valor crítico con el 5 % de significancia	decisión	valor crítico con el 10 % de significancia	decisión
1	22.10	0.00000	3.84	rechazar	2.71	rechazar
2	22.93	0.00001	5.99	rechazar	4.61	rechazar
3	23.19	0.00004	7.81	rechazar	6.25	rechazar
4	23.20	0.00012	9.49	rechazar	7.78	rechazar
5	26.29	0.00008	11.07	rechazar	9.24	rechazar
6	26.38	0.00019	12.59	rechazar	10.64	rechazar
7	26.63	0.00039	14.07	rechazar	12.02	rechazar
8	28.55	0.00038	15.51	rechazar	13.36	rechazar
9	29.96	0.00045	16.92	rechazar	14.68	rechazar
10	30.34	0.00075	18.31	rechazar	15.99	rechazar

Breusch-Godfrey

En el cuadro 9.12 se puede observar que la hipótesis nula de no autocorrelación es rechazada mediante la prueba LM para los primeros cinco grados autorregresivos.

Cuadro 9.12: Resultados de las pruebas de Breush-Godfrey

	Proceso Autorregresivo del Error				
	$AR(1)$	$AR(2)$	$AR(3)$	$AR(4)$	$AR(5)$
R^2	0.1279	0.1318	0.1351	0.1367	0.1641
LM	21.999***	22.406***	22.697***	22.692***	26.912***

(*) Rechaza la H_0 con un nivel de significancia: 10 %

(**) Rechaza la H_0 con un nivel de significancia: 5 %

(***) Rechaza la H_0 con un nivel de significancia: 1 %

Conclusión

Dado que los resultados de las pruebas de Rachas se contradicen con los resultados de la prueba de Durbin-Watson, Box-Pierce y Breush-Godfrey entonces se llega a la conclusión de que el problema de autocorrelación es de un orden superior a un AR(1), y por lo tanto no puede ser solucionado, esto se puede observar dado que el parámetro $\hat{\rho}$ es estadísticamente igual a 1 lo que sugiere la presencia de raíces unitarias. Dado que no se pudo solucionar el problema no tiene sentido interpretar los coeficientes estimados ya que estos son ineficientes.

Respuestas capítulo 7

1. Los resultados de la estimación se reportan en el cuadro 9.13.

Cuadro 9.13: Estimación del modelo MLP

	Variable dependiente $Admitido_i$ Estadísticos t entre paréntesis	
	MCO	MCP
Constante	-2.86733	0.65852
	(-3.42)***	7.348***
Qi	0.00313	0.00065
	(2.98)***	13.997***
Vi	0.00234	0.00045
	(3.44)***	8.658***
R^2	64.93	99.49
R^2 ajustado	57.92	98.97
F Global	9.26***	193.87***
No. de Obs.	13	5

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos cuadrados ordinarios

MCP: Mínimos cuadrados ponderados

El número de observaciones difiere, debido a que al transformar las series algunas de estas arrojaron datos indeterminados.

2. A pesar de que estadísticamente el modelo aparenta tener un buen ajuste, de todas formas el modelo MLP no es satisfactorio debido a los múltiples problemas que presenta, siendo estos principalmente la no normalidad en la distribución del término de error, heterocedasticidad, etc. Por lo tanto se estima el modelo por medio de los MCP, para eliminar el problema de heteroscedasticidad.

3. La estimación se reporta en el cuadro 9.14. Este modelo se puede expresar de la forma:

$$P(Admitidos = 1\,|\mathbf{x_i}) = \Phi(\beta^T \mathbf{x_i^T})$$

Donde $\Phi(z) = \int_{-\infty}^{z} \phi(v)dv = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{v}{2}} dv$, es decir, la función de densidad de la normal estándar, $\mathbf{x_i}$ corresponden al vector fila compuesto por algunas características de los estudiantes admitidos, y por consiguiente, el método de máxima verosimilitud para la estimación de los coeficientes viene definido por: $\underset{\hat{\beta}}{Max}\, P(Y\,|X) = \underset{\hat{\beta}}{Max} \left[\prod_{i=1}^{n} \Phi(\hat{\beta}^T \mathbf{x_i^T}) \right]$ equivalente a $\underset{\hat{\beta}}{Max}\, l(Y\,|X) = \underset{\hat{\beta}}{Max} \left[\prod_{i=1}^{n} \ln \left[\Phi(\hat{\beta}^T \mathbf{x_i^T}) \right] \right]$.

Los coeficientes estimados no son significativos conjunta e individualmente. Por su parte, la prueba de Wald, al tener un valor tan bajo, no nos permite rechazar la hipótesis nula de que todos los coeficientes asociados a variables explicativas son conjuntamente iguales a cero.

Cuadro 9.14: Estimación de modelos Logit y Probit

	Variable dependiente $Admitido_i$ Estadísticos t entre paréntesis			
	Probit EMV		Logit EMV	
	Coeficientes	Efecto marginal	Coeficientes	Efecto marginal
Constante	-45.1305553	–	-80.9004	–
	(-0.64)		(-0.72)	
Q_i	0.0461	0.0036	0.0834	0.0040
	(0.59)		(0.68)	
V_i	0.0266	0.0021	0.0469	0.0023
	(0.82)		(0.82)	
LRI	0.78834		0.78318	
Wald	0.88		0.96	
ln(L)	-1.89908		-1.94539	
No. de Obs.	13		13	

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

Wald: corresponde al estadístico de Wald para la significancia conjunta de todas las pendientes

EMV: Estimadores de Máxima Verosimilitud

4. Recuerden que la diferencia fundamental entre los modelos Logit y Probit es el supuesto sobre la distribución de los errores (normal en Probit, Logística en Logit). Así, tenemos:

$$P(Admitidos = 1\,|\mathbf{x_i}\,) = \Lambda(\beta^T \mathbf{x_i^T})$$

Donde $\Lambda(z) = \frac{e^z}{1+e^z}$, es decir, la función de densidad logística, $\mathbf{x_i}$ corresponden al vector fila compuesto por las mismas características de los estudiantes admitidos **i** que fueron consideradas en el ejercicio anterior, y por consiguiente, el método de máxima verosimilitud para la estimación de los coeficientes viene definido por: $\underset{\hat{\beta}}{Max}\, P(Y\,|X\,) = \underset{\hat{\beta}}{Max} \left[\prod_{i=1}^{n} \Lambda(\hat{\beta}^T \mathbf{x_i^T})\right]$ equivalente a $\underset{\hat{\beta}}{Max}\, l(Y\,|X\,) = \underset{\hat{\beta}}{Max} \left[\prod_{i=1}^{n} \ln\left[\Lambda(\hat{\beta}^T \mathbf{x_i^T})\right]\right]$.

Los resultados se muestran en el cuadro 9.14.

Los coeficientes estimados no son significativos conjunta e individualmente. Por su parte, la prueba de Wald, al tener un valor tan bajo, no nos permite rechazar la hipótesis nula de que todos los coeficientes asociados a variables explicativas son conjuntamente iguales a cero.

5. A continuación se presentan los cálculos de los efectos marginales para el modelo Probit y Logit:

Cuadro 9.15: Efectos marginales modelo Probit

Modelo Probit				
Coeficientes Probit		0.0461441	0.0265665	-45.1305553
(i)	(b'x(i)')	Q(i)	V(i)	
1	4.5505357	5.86782E-07	3.37828E-07	
2	-8.1458203	7.18379E-17	4.13592E-17	
3	-3.4055233	5.5801E-05	3.21262E-05	
4	4.3686507	1.32053E-06	7.60266E-07	
5	-9.2505873	4.81955E-21	2.77475E-21	
6	0.0060147	0.018408499	0.010598308	
7	5.0679747	4.87233E-08	2.80514E-08	
8	2.9283297	0.000252909	0.000145607	
9	-3.6017333	2.806E-05	1.61549E-05	
10	-1.5736503	0.005336889	0.003072602	
11	-1.4620883	0.006321645	0.003639555	
12	-0.3994283	0.016997378	0.009785885	
13	9.7240577	5.39722E-23	3.10733E-23	
Media del efecto marginal		0.003646395	0.002099336	

Cuadro 9.16: Efectos marginales modelo Logit

Modelo Logit					
Coeficientes Logit			0.0833849	0.0468906	-80.9004337
(i)	(b'x(i)')	L(b'x(i)')	Q(i)	V(i)	
1	8.2619203	0.999741904	2.15158E-05	1.20992E-05	
2	-14.4577837	5.26095E-07	4.38683E-08	2.46689E-08	
3	-5.8583137	0.002847922	0.000236797	0.00013316	
4	7.8439233	0.999608026	3.26719E-05	1.83727E-05	
5	-16.5434787	6.53519E-08	5.44936E-09	3.06439E-09	
6	0.0273933	0.506847897	0.020842315	0.011720451	
7	9.2527863	0.999904165	7.99043E-06	4.49333E-06	
8	5.1853593	0.994433237	0.0004616	0.000259576	
9	-6.5924897	0.001368747	0.000113977	6.40935E-05	
10	-2.4698637	0.077998036	0.00599657	0.003372107	
11	-2.5250627	0.074119762	0.005722374	0.003217915	
12	-0.6494387	0.343116036	0.018793908	0.010568551	
13	17.4321143	0.999999973	2.24087E-09	1.26013E-09	
Media del efecto marginal			0.004017675	0.002259296	

Respuestas capítulo 8

1. Las variables endógenas son: C_t, I_t, T_t, M_t, Y_t. Las variables exógenas son: X_t, G_t, Y_{t-1}, C_{t-1}, M_{t-1}

Cuadro 9.17: Resumen de las ecuaciones

Ecuación	Variables Endógenas Incluidas (gi)	Variables Exógenas Excluidas (ki)	Condición de Orden ki≥gi-1	Identificación	Método de Estimación
8.23	3	4	4>2	Sobre	MC2E
8.24	1	4	4>0	Sobre	MC2E
8.25	2	5	5>1	Sobre	MC2E
8.26	2	3	3>1	Sobre	MC2E

La condición de orden para la ecuación 8.24 no es relevante ya que como ésta no se encuentra en función de ninguna variable endógena, no presenta problema de sesgo por simultaneidad. De acuerdo con esto, la ecuación 8.24 se puede estimar por medio de MCO.

2. Según los resultados del punto anterior, el método adecuado para estimar las ecuaciones estructurales 8.23, 8.25 y 8.26 es el de MC2E. La ecuación 8.24 se debe estimar por medio de MCO. Cabe mencionar que la ecuación 8.27 es una identidad macroeconómica donde no existe ningún parámetro y por lo tanto no se debe estimar. Los resultados de la estimación se reportan en el cuadro 9.18.

Cuadro 9.18: Resultados de las estimaciones

	Estadísticos t entre paréntesis			
	Ecuación 8.23 1970-1997 MC2E	Ecuación 8.24 1970-1997 MCO	Ecuación 8.25 1970-1997 MC2E	Ecuación 8.26 1970-1997 MC2E
Variable dependiente	C_t	I_t	T_t	M_t
Constante	1363613.99 (1.18)	202377.23 (0.23)	-12216524.21 (-17.91)***	-1855464.41 (-2.61)**
Y_t	0.299 (3.41)***	–	0.475 (23.19)**	0.652 (4.55)***
Y_{t-1}	–	0.231 (8.47)***	–	-0.591 (-4.28)***
C_t	–	–	–	–
C_{t-1}	0.481 (4.89)***	–	–	–
M_{t-1}	–	–	–	0.974 (16.75)***
T_t	0.011 (0.12)	–	–	–
R^2	0.9954	0.7418	0.955	0.9952
R^2 ajustado	0.9948	0.7314	0.9537	0.9945
Wald	4994.76***	n.a	n.a	4713.38***
No. de Obs.	27	27	27	27

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos cuadrados ordinarios

MC2E: Mínimos cuadrados en dos etapas

3. **Ecuación 8.23**:
$\hat{\alpha}_1 = 1363613.99$: Es la parte de la función de consumo privado nacional que no depende del PIB, ni del Consumo privado nacional del periodo pasado, ni de la recaudación directa.
$\hat{\alpha}_2 = 0.299$: Un incremento de un millón de moneda local en el PIB incrementará la función de consumo privado en 0.299 millones de moneda local constantes en 1986.
$\hat{\alpha}_3 = 0.481$: Un incremento de un millón de moneda local en el consumo privado nacional del periodo anterior incrementará la función de Consumo privado nacional en 0.481 millones de moneda local constantes en 1986.
$\hat{\alpha}_4 = 0.011$: Un incremento de un millón de moneda local en la recaudación directa incrementará la función de consumo privado en 0.011 millones de moneda local constantes en 1986.

Ecuación 8.24:
$\hat{\beta}_1 = 202377.23$: La parte de la función de inversión privada que no depende del PIB del periodo anterior es igual a 202,337.23 millones de moneda local (Constantes de 1986).
$\hat{\beta}_2 = 0.231$: Un incremento de un millón de moneda local en el PIB del periodo anterior incre-

mentará la función de Inversión privada en 0.231 millones de moneda local constantes en 1986.

Ecuación 8.25:
$\hat{\gamma}_1$ = -12216524.21: La parte de la función de recaudación directa que no depende del PIB es igual a -12,216,524.21 millones de moneda local constantes en 1986.
$\hat{\gamma}_2$ = 0.475: Un incremento de un millón de moneda local en el PIB incrementará la función de recaudación directa en 0.475 millones de moneda local constantes en 1986.

Ecuación 8.26:
$\hat{\delta}_1$ = -1855464.41: La parte de la función de importaciones que no dependen del PIB, ni del PIB del periodo anterior, ni de las importaciones del periodo anterior tienen un valor de - 1,855,464.41 millones de moneda local constantes de 1986.
$\hat{\delta}_2$ = 0.652: Un incremento de un millón de moneda local en el PIB incrementará la función de importaciones en 0.652 millones de moneda local constantes de 1986.
$\hat{\delta}_3$ = -0.591: Un incremento de un millón de moneda local en el PIB del periodo anterior reducirá la función de importaciones en 0.591 millones de moneda local constantes de 1986.
$\hat{\delta}_4$ = 0.974: Un incremento de un millón de moneda local en las importaciones del periodo anterior incrementará la función de importaciones en 0.974 millones de moneda local constantes de 1986.

4. Para determinar el efecto de un incremento en el gasto público se debe emplear la forma reducida del sistema de ecuaciones:

$$C_t = \pi_{11} + \pi_{12}Y_{t-1} + \pi_{13}C_{t-1} + \pi_{14}M_{t-1} + \pi_{15}X_t + \pi_{16}G_t + \mu_{1t} \tag{9.6}$$

$$I_t = \pi_{21} + \pi_{22}Y_{t-1} + \pi_{23}C_{t-1} + \pi_{24}M_{t-1} + \pi_{25}X_t + \pi_{26}G_t + \mu_{2t} \tag{9.7}$$

$$Y_t = \pi_{31} + \pi_{32}Y_{t-1} + \pi_{33}C_{t-1} + \pi_{34}M_{t-1} + \pi_{35}X_t + \pi_{36}G_t + \mu_{3t} \tag{9.8}$$

$$M_t = \pi_{41} + \pi_{42}Y_{t-1} + \pi_{43}C_{t-1} + \pi_{44}M_{t-1} + \pi_{45}X_t + \pi_{46}G_t + \mu_{4t} \tag{9.9}$$

$$T_t = \pi_{51} + \pi_{52}Y_{t-1} + \pi_{53}C_{t-1} + \pi_{54}M_{t-1} + \pi_{55}X_t + \pi_{56}G_t + \mu_{5t} \tag{9.10}$$

En este caso, no es necesario estimar las ecuaciones 9.9 y 9.10. Los resultados de las estimaciones de las ecuaciones reducidas se presentan en el cuadro 9.19.

Cuadro 9.19: Resultados de las estimaciones

	Estadísticos t entre paréntesis		
	Ecuación 9.6	Ecuación 9.7	Ecuación 9.8
	1970-1997	1970-1997	1970-1997
	MCO	MCO	MCO
Variable dependiente	C_t	I_t	T_t
Constante	2112132.47	2643478.03	4190022.52
	(1.54)	(1.26)	(2.13)**
$Yt-1$	0.723	1.223	1.545
	(3.30)***	(3.66)***	(4.92)***
C_{t-1}	-0.161	-1.566	-1.078
	(-0.61)	(-3.92)***	(-2.87)***
$Mt-1$	0.192	0.935	0.127
	(1.91)*	(6.12)***	(0.88)
X_t	-0.188	-0.805	-0.039
	(-2.06)*	(-5.80)***	(-0.30)
G_t	-0.173	-0.5969	0.1924
	(-0.50)	(-1.14)	(0.39)
R^2	0.9926	0.9162	0.9947
R^2 ajustado	0.9909	0.8962	0.9934
F Global	566.32***	45.89***	783.39***
No. de Obs.	27	27	27

(*) nivel de significancia: 10 %

(**) nivel de significancia: 5 %

(***) nivel de significancia: 1 %

MCO: Mínimos cuadrados ordinarios

Se observa que en las tres ecuaciones estimadas el gasto público no es significativo individualmente. De acuerdo con esto, el gasto público no tiene efecto sobre el consumo, la inversión o el PIB.

Bibliografía

J. Alonso y M. Cárdenas. Tasa de cambio real en colombia. En T. M. Editores, editor, *Tasa de Cambio en Colombia*. Fedesarrollo, 1997.

J. Alonso, I. Fainboim, y M. Olivera. La sostenibilidad de la política fiscal colombiana. *Coyuntura Económica, Fedesarrollo*, 27:101118, 1997.

J. Alonso y F. Romero. The day of the week efect: The colombian exchange rate and stock market case. 2008.

T. Amemiya. Qualitative response models: A survey. *Journal of Economic Literature*, 19(4):1483–1536, 1981.

D. A. Belsley, E. Kuh, y R. E. Welsch. Regression diagnostics : identifying influential data and sources of collinearity. *Wiley series in probability and mathematical statistics*, 1980.

L. F. Bernat. Diferencias salariales por género en las siete principales áreas metropolitanas colombianas. ¿evidencia de discriminación? *Investigaciones sobre género y desarrollo en Colombia*, (1), 2004.

H. Bohn. The sustainability of budget deficits in a stochastic economy. *Journal of Money, Credit and Banking*, 27(1):257–71, 1995.

G. E. P. Box y D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509–1526, 1970.

T. S. Breusch. Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17 (31):334–55, 1978.

T. S. Breusch y A. R. Pagan. A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47(5):1287–94, 1979.

J. Durbin. Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society*, 22(1):139–153, 1960.

J. Durbin. Testing for serial correlation in least-squares regression when some of the regressors are lagged dependent variables. *Econometrica*, 38(3):410–21, 1970.

J. Durbin y G. S. Watson. Testing for serial correlation in least squares regression. i. *Biometrika*, 37 (3-4):409–428, 1950.

J. Durbin y G. S. Watson. Testing for serial correlation in least squares regression. ii. *Biometrika*, 38 (1-2):159–178, 1951.

J. Durbin y G. S. Watson. Testing for serial correlation in least squares regression.iii. *Biometrika*, 58 (1):1–19, 1971.

L. G. Godfrey. Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica*, 46(6):1293–1301, 1978.

S. M. Goldfeld y R. E. Quandt. Some tests for homoscedasticity. *Journal of the American Statistical Association*, 60(310):539–547, 1965.

W. H. Greene. *Econometric analysis*. Prentice Hall, Upper Saddle River, N.J., 5th edición, 2003.

A. Greiner, U. Koeller, y W. Semmler. Testing sustainability of german fiscal policy. evidence for the period 1960 â 2003. CESifo Working Paper Series CESifo Working Paper No. 1386, CESifo GmbH, 2005.

G. M. Ljung y G. E. P. Box. The likelihood function of stationary autoregressive-moving average models. *Biometrika*, 66(2):265–270, 1979.

J. A. Mincer. *Schooling, Experience, and Earnings*. Number minc74-1 en NBER Books. National Bureau of Economic Research, Inc, 1974.

R. Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14 (3):693–709, 1973.

D. Romer. Openness and inflation: Theory and evidence. *The Quarterly Journal of Economics*, 108 (4):869–903, 1993.

F. S. Swed y C. Eisenhart. Tables for testing randomness of grouping in a sequence of alternatives. *The Annals of Mathematical Statistics*, 14(1):66–87, 1943.

A. P. Thirlwall. The balance of payments constraint as an explanation of internacional growth rate differences. *Banca Nazionale del Lavoro Quarterly Review*, 1979.

A. Wald y J. Wolfowitz. On a test whether two samples are from the same population. *The Annals of Mathematical Statistics*, 11(2):147–162, 1940.

H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–38, 1980.

Índice alfabético

EasyReg: Aplicaciones para un curso de econometría

El análisis empírico de la realidad a la luz de la teoría es cada día más una parte fundamental en la formación de un economista. La microeconomía, la macroeconomía y todo el andamiaje teórico de los programas de pregrado son complementados con la econometría para constituir la "caja de herramientas" de un economista. Así la econometría no se puede entender como un fin, sino como un medio para encontrar las regularidades presentes en una realidad.

Este libro pretende cumplir dos funciones. La primera es convertirse en un documento de consulta para economistas que desean emplear métodos econométricos convencionales y una herramienta gratuita como EasyReg. La segunda función es complementar un libro de texto convencional de econometría brindando el puente entre la teoría econométrica y la práctica econométrica mediante el uso del paquete econométrico gratuito EasyReg. Así, este libro no pretende ser un sustituto de los manuales de econometría actualmente disponibles en el mercado, sino un complemento a estos.

ISBN: 978-958-8357-16-4

UNIVERSIDAD ICESI