Econometría
Notas de Clases

Rodrigo Montero P.*

20 de mayo de 2013

1. Introducción

En términos semánticos econometría significa medición de la economía, lo cual por cierto no es una tarea fácil. A diferencia de otras ciencias, la economía no es una ciencia exacta, es más bien una ciencia social, por lo tanto, realizar mediciones en este campo constituye todo un desafío. No obstante, el esfuerzo bien vale la pena. En efecto, la econometría es útil al menos por tres razones: en primer lugar, permite conocer la forma en que se relacionan las variables en el mundo real, identificando ciertas regularidades que a su vez posibiliten sistematizar ciertos hechos. Por ejemplo, si una determinada teoría establece que si aumenta la tasa de interés (r), entonces, el consumo (C) disminuye, la econometría permite corroborar aquello proporcionando un fundamento empírico, lo cual se logra mediante el análisis de los datos. Además permite comprender cuáles son los órdenes de magnitud involucrados en el fenómeno. En segundo lugar, la econometría es útil por cuanto permite realizar proyecciones respecto de lo que sucedería frente a cierto fenómeno. Siguiendo con el mismo ejemplo anterior, si el trabajo empírico permite identificar la existencia de una relación inversa entre la tasa de interés y el nivel de consumo agregado, entonces, la econometría permitirá proyectar qué sucederá frente a un aumento de la tasa de política monetaria (TPM). Esto es de mucha utilidad, pues si el objetivo consiste en desacelerar la demanda agregada para mantener bajo control la inflación, entonces, el Banco Central cuenta con una herramienta que ha sido validada por el análisis empírico. Así, el trabajo econométrico permite a todos los agentes económicos aprender del mundo real, y como consecuencia, tomar mejores decisiones.

*Universidad Diego Portales. Facultad de Economía y Empresa. Cualquier comentario o sugerencia enviarlo a rodrigo.montero@udp.cl. Se agradece particularmente el financiamiento y apoyo provisto por la Universidad Diego Portales a través del Fondo para el Desarrollo de Proyectos Docentes. Finalmente, cabe destacar que para la elaboración de este documento fue de vital importancia la ayuda de Javier Beth, destacado alumno de la carrera de ingeniería comercial de la Universidad Diego Portales.

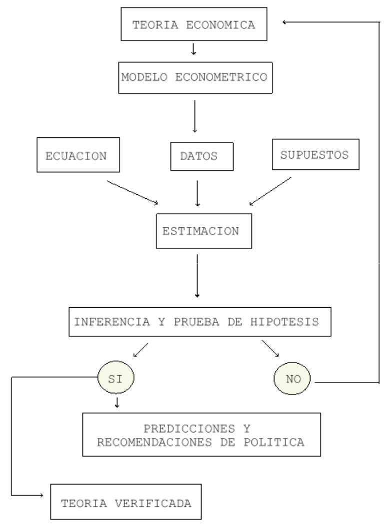
Para un gerente de producción sería muy útil contar con una estimación confiable de cuál será la demanda para el próximo año del bien que produce, ya que esto le permitirá decidir de mejor manera cuánto es lo que deberá producir, y por ende, si será necesario o no adquirir nueva maquinaria, o incluso, contratar más mano de obra. Por otro lado, para una empresa constructora sería muy beneficioso saber en cuánto podría crecer la demanda por viviendas nuevas, antes de decidir si debe o no construir un nuevo proyecto inmobiliario. En cambio, para un inversionista sería muy útil saber cuánto se podría depreciar (o apreciar) el peso chileno respecto al dólar (o el euro), con el objetivo de poder reestructurar a tiempo su portafolio.

Tal como se ha podido apreciar en estos ejemplos, el trabajo econométrico permite descubrir ciertas regularidades empíricas que motivan avances teóricos.[1]

Para poder cumplir con los objetivos anteriores, lo que hace la econometría es aplicar métodos matemáticos y estadísticos al análisis de datos económicos. Frente a esto, se debe hacer una salvedad, muy importante: aún cuando la econometría utilice técnicas estadísticas muy sofisticadas para adquirir un mayor entendimiento del mundo real, lo más importante es el modelo teórico subyacente. En efecto, la econometría permite hacer un pronunciamiento respecto de la validez de un modelo, pero difícilmente podrá establecer relaciones causales que den origen a modelos teóricos. Así, la econometría plantea una relación estadística a priori, lo que no implica que exista una cierta relación de causalidad. En otras palabras, no se puede hacer teoría a partir de los datos. Una forma esquemática respecto de cómo es posible utilizar la econometría en el análisis empírico se presenta en la figura 1.

Figura 1: El rol de la econometría

[1] Por ejemplo, suponga que se tiene una determinada hipótesis, que establece que los trabajadores más productivos debieran percibir mayores salarios. Luego, el trabajo econométrico permitirá contrastar esto con lo que dicen los datos, es decir, la relación empírica existente entre el salario y los años de escolaridad.



Primero se requiere de un modelo económico, el cual se plantea en términos de un modelo econométrico, es decir, susceptible de ser estimado. A continuación, con una (o varias) ecuación (es), datos y supuestos, es posible llevar a cabo la estimación. La inferencia y las pruebas de hipótesis permiten analizar la teoría a la luz de los resultados empíricos. Si se valida la teoría se pueden hacer recomendaciones de política, mientras que un rechazo de la misma pudiera dar cabida a revisar la teoría.

Cabe mencionar que uno de los grandes hitos de la econometría ocurrió en el año 1930 con la fundación de la Sociedad Econométrica y del journal *Econometrica*. A partir de entonces, los principales avances en esta materia han sido publicados en esta prestigiosa revista. Una excelente descripción acerca de la evolución que ha tenido la econometría desde sus inicios se puede encontrar en Pesarán (1987). Adicionalmente, para una revisión de los artículos más importantes en el desarrollo de la econometría se puede ver en Hendry y Morgan (1995).

Ahora bien, si la econometría consiste en la aplicación de técnicas estadísticas, entonces, cabe preguntarse qué distingue a un economista de un estadístico. Básicamente, es la preocupación por la violación de los supuestos estadísticos que validan el modelo, ya que los datos disponibles rara vez cumplen con las propiedades deseadas. En la práctica, sin embargo, el prestigio del econometrista parece depender cada vez más de la aplicación en sí de las técnicas de estimación, pasando de esta manera a un segundo plano los datos que se emplearán para llevar a cabo la estimación. No obstante, una buena estimación depende tanto de la calidad de los datos utilizados como del método de estimación utilizado.

Con el objetivo de introducir adecuadamente al alumno en el estudio de la econometría, este documento de ha organizado de la siguiente manera. Luego de esta breve introducción, la sección 2 desarrolla el modelo clásico de regresión lineal y plantea sus principales supuestos. La sección 3 presenta el estimador de mínimos cuadrados ordinarios (MCO). La sección 4 analiza las consecuencias que tiene para la estimación por MCO la violación de los supuestoa del modelo clásico de regresión lineal. Finalmente, en la sección 5 se analizan las dificultades que surgen en la práctica cuando los datos presentan problemas.

2. El modelo clásico de regresión lineal

Un estudio econométrico comienza con un conjunto de premisas en torno a alguna pregunta que se quiera responder. La teoría económica establece la relación existente entre un conjunto de variables (exógenas) y una variable endógena, y el objetivo de la econometría consiste en estimar dicha relación. En términos más concretos, la regresión corresponde a un estudio de dependencia entre una variable dependiente (Y) y una o más variables explicativas (X). El análisis de regresión tiene por objetivo estimar y/o predecir el promedio poblacional de la variable dependiente para valores fijos de la(s) variable(s) explicativa(s), es decir: $E(Y|X)$. En econometría interesa la dependencia estadística entre las variables, las cuales son aleatorias, es decir, variables que tienen una distribución de probabilidad. En cambio, la dependencia determinística trata relaciones que son exactas, en donde no hay espacio para la econometría. Un ejemplo de una relación determinística podría ser aquella que establece que la fuerza de un cuerpo (F) es igual a su masa (m) multiplicada por la aceleración (a): $F = ma$. La regresión es una relación estadística que no implica causalidad a priori (no se puede hacer teoría a partir de los datos). El trabajo empírico a menudo descubre regularidades empíricas que motivan avances teóricos. Por ejemplo, se tiene una hipótesis/teoría (trabajadores más productivos reciben mayores salarios), luego ésta se contrasta con lo que dicen los datos (relación entre salario y años de escolaridad), y si se encuentra una regularidad se obtiene una teoría respaldada por los datos. A modo de ejemplo, considere las siguientes relaciones entre dos variables:

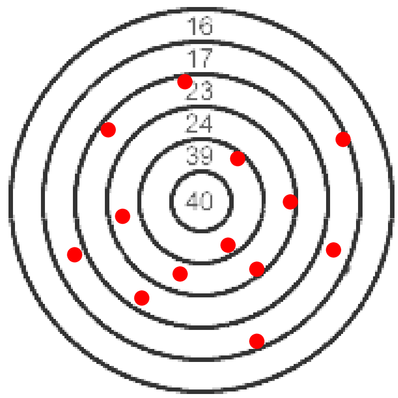
- Ventas y gasto en publicidad
- Salarios y capital humano
- Tipo de cambio nominal y (*spread*) de tasas de interés doméstica y externa
- Inversión y utilidades
- Precio del cobre e inventarios
- Tasa de desempleo y producto interno bruto
- Consumo e ingreso

Luego, a través de las técnicas econométricas es posible encontrar estimaciones para los parámetros poblacionales (que son desconocidos) del modelo (elasticidades, semi elasticidades, sensibilidades, etc.). En este contexto, el modelo de regresión lineal múltiple se utiliza para estudiar la relación existente entre una variable dependiente, y un conjunto de variables independientes, que se asume tiene la siguiente forma:

$$y = \beta_1 + \beta_2 X_2 + \cdots + \beta_k X_k + u \tag{1}$$

donde y representa la variable dependiente o explicada, $X_2, X_3, \ldots, X_k$ son las variables independientes o explicativas (también llamadas regresores), y finalmente, u es un shock aleatorio no observable que afecta la relación teórica entre las variables, haciendo que esta no sea determinística. En otras palabras, u es una variable estocástica y captura aquella parte de y que no puede ser explicada por X.[2] El apellido estocástico que se le agrega al término de error viene del griego *stokhos*, que significa tiro al blanco (*target*). Por lo tanto, el término de error captura el tamaño de los errores que se cometen al estimar la relación existente entre y y X.

Figura 2: El shock estocástico



Así, caracterizar el término de error consiste en describir la naturaleza de las fallas cometidas: ¿se distribuyen uniformemente en torno al centro?, ¿cuál es la distancia promedio de la falla respecto del centro?

Este error (perturbación) que altera la relación entre las variables y que hace que ésta no sea determinística surge por diversas razones. En primer lugar, no es posible capturar todos los factores que influyen sobre una determinada variable económica en un modelo, es decir, siempre se van a omitir algunos factores, los cuales serán capturados por el término de error (u). Una segunda razón que da origen a esta perturbación es la existencia de errores de medida en las variables contenidas en el modelo. Finalmente, como la economía es una ciencia social, no exacta, está sujeta al comportamiento de los seres humanos. En efecto, el comportamiento humano es de tal forma que, bajo circunstancias idénticas, las respuestas frente a una determinada situación diferirán de una manera aleatoria, lo cual es razonable de suponer. Es decir,

[2] En este caso particular, se ha asumido que el término de error (shock) es aditivo, sin embargo, podría no serlo. Por ejemplo, la relación entre y y X podría tener la siguiente forma:

$$y = aX^{\alpha}exp(u)$$

donde *exp* corresponde a la constante de Neper (número de euler) que es igual a 2,7182. Más claramente:

$$exp(u) = e^u$$

frente a un mismo estímulo se obtienen respuestas diversas. El término de error puede representar de buena forma esta aleatoriedad inherente al comportamiento humano.

La función de consumo keynesiana constituye un buen caso para ejemplificar la idoneidad de incorporar un término de error. En efecto, esta relación establece que el consumo (C) depende linealmente del ingreso disponible (YD) de la siguiente manera:

$$C = \delta_1 + \delta_2 YD + u \tag{2}$$

donde se ha incluido un término de error (u), reflejando el hecho que se trata de una relación sujeta a eventos aleatorios. Por ejemplo, el consumo de una persona pudiera verse afectado por su estado anímico en un día particular, o bien, porque ha comenzado una etapa de ahorro con el objetivo de adquirir un bien durable en el futuro. Pudiera ser que el consumo aumentara debido a que logró obtener financiamiento por otros medios, o bien, porque espera una notoria mejoría de su condición económica en el futuro. Por lo tanto, lo que la teoría establece es que $\delta_1 > 0$, es decir, que hay un cierto nivel de consumo mínimo, de subsistencia, denominado consumo autónomo, y que además, $\delta_2 \in [0, 1]$, es decir, que por cada peso adicional, el individuo puede incrementar su consumo en una cuantía no superior a ese peso. Luego, el análisis econométrico permitirá corroborar estas hipótesis al proporcionar estimaciones para dichos paramétros, utilizando la información que exista de consumo e ingreso disponible para un conjunto de personas.

Otro ejemplo lo constituye la relación existente entre el salario que perciben las personas (y) y su nivel de escolaridad (S). La evidencia empírica muestra que personas con mayor escolaridad (capital humano) reciben también un mayor salario, por lo tanto, se puede plantear el siguiente modelo:

$$y = \gamma_1 + \gamma_2 S + u \tag{3}$$

La teoría señala que $\gamma_2 > 0$, es decir, que existe un premio o retorno asociado a la inversión en capital humano. En realidad, también es posible apreciar un premio salarial (decreciente) por la experiencia laboral del individuo, la cual puede aproximarse a través de la edad de la persona (E). En este contexto, el modelo puede ampliarse para quedar de la siguiente forma:

$$y = \gamma_1 + \gamma_2 S + \gamma_3 E + \gamma_4 E^2 + u \tag{4}$$

Ahora, el shock estocástico (u) refleja las diferencias salariales existentes entre personas con igual capital humano y edad. Éstas podrían darse por diferencias en factores no observables, como por ejemplo, las habilidades de los trabajadores o su poder de negociación salarial.

En términos más generales, es posible señalar que el shock estocástico tiene su justificación en tres razones:

1. Por omisión de factores que afectan la relación entre las variables, tal como se mencionó en los ejemplos anteriores. El término de error representa todos esos elementos no considerados explícitamente en la ecuación, pero que tienen un impacto moderado y causas independientes.

2. Porque las variables están medidas con error. Por ejemplo, la variable dependiente (salario) podría estar medida con error, de forma tal que en lugar de observar la variable verdadera (Y^*), se observa a la variable con ruido: $Y = Y^* + e$. Ese error e es una razón para incorporar un término de error (shock) en el modelo.

3. por el comportamiento humano; algunos creen que el comportamiento humano es de tal forma que bajo circunstancias idénticas las acciones llevadas a cabo diferirán de una manera aleatoria. Así, frente a un mismo estímulo se obtienen respuestas diversas. Pues bien, el término de error puede representar esta aleatoriedad inherente en el comportamiento humano.

Note que para el planteamiento de los modelos considerados hasta ahora se han realizado una serie de supuestos, los cuales van a hacerse explícitos a continuación.

2.1. Supuestos del modelo clásico de regresión lineal

El modelo de regresión lineal clásico se construye sobre la base de un conjunto de supuestos acerca de cómo los datos reales (que el investigador observa) se generan a partir de un proceso generador de datos. A continuación se explicitan uno a uno estos supuestos.

Supuesto 1: El modelo es lineal en los parámetros

Suponga que usted dispone de un conjunto de observaciones (n) para las cuales se cumple la siguiente relación:

$$\begin{aligned} y_1 &= \beta_1 + \beta_2 x_{12} + \cdots + \beta_k x_{1k} + u_1 \\ y_2 &= \beta_1 + \beta_2 x_{22} + \cdots + \beta_k x_{2k} + u_2 \\ &\vdots \\ y_n &= \beta_1 + \beta_2 x_{n2} + \cdots + \beta_k x_{nk} + u_n \end{aligned}$$

Estas n ecuaciones anteriores pueden agruparse y presentarse en forma matricial, de la siguiente manera:

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} x_{12} \\ \vdots \\ x_{n2} \end{pmatrix} + \cdots + \beta_k \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix} + \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

Finalmente, lo anterior puede resumirse de la siguiente manera:

$$y = X\beta + u \tag{5}$$

donde:

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad X = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1k} \\ 1 & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & \cdots & x_{nk} \end{pmatrix} \quad y \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

Los vectores columna y y u tienen dimensión n, mientras que la matriz X es de nxk.[3] Si el modelo es lineal entonces puede expresarse como $y = X\beta + u$.[4] La linealidad se refiere a la manera en que el vector de parámetros y el shock estocástico entran en la ecuación y no a la relación que existe entre las variables. En efecto, considere el modelo log lineal:

$$y = exp(\beta_1)x_2^{\beta_2}x_3^{\beta_3}\cdots x_k^{\beta_k}exp(u) = exp(\beta_1)\prod_{j=2}^{k}x_j^{\beta_j}exp(u)$$

Aplicando logaritmo a la expresión anterior:

$$ln(y) = \beta_1 + \beta_2 ln(x_2) + \cdots + \beta_k ln(x_k) + u$$

Luego, el modelo es lineal, desde el punto de vista de la definición que acá se maneja. Cabe señalar que a esta ecuación corresponde a una especificación de elasticidad constante (η). En efecto, recuerde que la elasticidad de la variable y con respecto a la variable independiente x_k se define como:

$$\eta_k = \frac{dy/y}{dx_k/dx_k} = \frac{\partial ln(y)}{\partial ln(x_k)} = \beta_k$$

Por lo tanto, se puede apreciar que la elasticidad es constante.

Supuesto 2: Rango completo y condición de identificación
Se asume que no existe una relación lineal exacta entre los regresores (x). Es decir,

---

[3]Para indicar una observación particular del modelo matricial $y = X\beta + u$, por ejemplo, la i-ésima, se recurrirá a la siguiente notación:

$$y_i = x_i'\beta + u_i$$

donde x_i' es el i-ésimo vector fila de la matriz X.

[4]Se debe tener cuidado en este aspecto, ya que si bien el modelo pudiera ser no lineal, podría ser linealizable. En efecto, considere el siguiente modelo:

$$y = ax^{\beta}exp(u)$$

Note que este modelo no lineal, es intrínsicamente lineal ya que tomando logaritmo de la ecuación anterior se llega a la siguiente expresión:

$$ln(y) = ln(a) + \beta ln(x) + u$$

Por lo tanto, este modelo puede escribirse como:

$$y^* = \beta_1 + \beta_2 x^* + u$$

Por otro lado, note que el modelo:

$$y = ax^{\beta} + u$$

no es lineal en los parámetros y tampoco es linealizable.

x es una matriz de dimensión nxk con rango k.[5] En otras palabras, X tiene rango columna completo si las columnas de X son linealmente independientes. Por otro lado, la condición de identificación exige que existan a lo menos $k+1$ observaciones, es decir, $n > k+1$.

Supuesto 3: El shock estocástico tiene media nula

El tercer supuesto establece que el valor esperado del error condicional en X es cero, para cada una de las observaciones. Por lo tanto:[6]

$$E(u|X) = \begin{pmatrix} E(u_1|X) \\ E(u_2|X) \\ \vdots \\ E(u_n|X) \end{pmatrix} = 0 \tag{6}$$

De la misma manera, se asume lo siguiente:

$$E(u_i|u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n) = 0 \tag{7}$$

Es decir, se ha asumido que los errores son puramente fenómenos aleatorios. Por lo tanto, note que este tercer supuesto tiene la siguiente implicancia:

$$E(y|X) = X\beta + E(u|X) = X\beta \tag{8}$$

Estos primeros tres supuestos que se han enunciado constituyen el modelo de regresión lineal. Luego, la regresión de y en contra de X es la media de y condicional en X, matemáticamente $E(y|X)$.

Supuesto 4: Perturbaciones esféricas

El cuarto supuesto se refiere a las varianzas y covarianzas de los errores. De esta manera, se asume lo siguiente:[7]

$$var(u_i|X) = E(u_i^2|X) = \sigma^2 \quad \forall i = 1, \ldots, n \tag{9}$$

[5] El rango columna de una matriz es la dimensión del espacio vectorial que es barrido por sus columnas. Por lo tanto, el rango columna de una matriz corresponde al número de vectores columna linealmente independientes que contenga dicha matriz. A modo de ejemplo, considere la matriz A:

$$A = \begin{pmatrix} 1 & 5 & 6 \\ 2 & 6 & 8 \\ 7 & 1 & 8 \end{pmatrix}$$

Dado que la tercera columna de esta matriz corresponde a la suma de las dos primeras, entonces, el rango columna de A es dos.

[6] Si además se asumiera que los regresores son no estocásticos, entonces, este supuesto podría plantearse como: $E(u) = E(u|X) = 0$. En términos concretos asumir que los regresores son no estocásticos no tiene ninguna importancia fundamental para los resultados estadísticos que se presentarám más adelante.

[7] Si se asume que los regresores son no estocásticos, entonces, los resultados son incondicionales en X.

$$cov(u_i, u_j|X) = E(u_iu_j|X) - E(u_i|X)E(u_j|X) = E(u_iu_j|X) = 0 \quad \forall i \neq j \tag{10}$$

Cuando la varianza es constante se dice que es homocedástica. Por otro lado, cuando los errores no están correlacionados entre sí, se dice que no están autocorrelacionados.[8]

En conjunto, los supuestos 3 y 4 se resumen de la siguiente manera en la matriz de varianzas y covarianzas:

$$E(uu') = \begin{pmatrix} E(u_1u_1) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2u_2) & \cdots & E(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_nu_n) \end{pmatrix}$$

Así:

$$E(uu') = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{11}$$

Es decir:

$$E(uu') = \sigma^2 I_n \tag{12}$$

que corresponde a la forma que tendrá la matriz de varianzas y covarianzas del término de error cuando éstos sean homocedásticos y no autocorrelacionados, es decir, esféricos.

Supuesto 5: Regresores no estocásticos

Opcionalmente se puede asumir que la matriz X es no estocástica. Este supuesto se realiza sólo por conveniencia matemática. En efecto, con este supuesto es posible utilizar los resultados elementales en estadística para obtener las propiedades de interés. El supuesto establece que X es una matriz de constantes conocidas en la distribución de probabilidad de y. Con esta simplificación, los supuestos 3 y 4 pueden formularse incondicionalmente, como $E(u_i) = 0$, y $E(u_iu_j) = 0$.

Una visión alernativa a la anterior, pero con consecuencias equivalentes, consiste en considerar a X fijo pero en repetidas muestras, lo que es equivalente a realizar el análisis estadístico condicional a la muestra que se ha observado. De esta manera, el modelo de regresión y sus supuestos aplican sólo al conjunto particular de datos (X) que se han observado. En todo caso, cualquiera sea la estrategia a seguir esto permitirá desligarse del origen de la variación en X y concentrarse sólo en la relación existente entre X e y.[9]

Supuesto 6: Normalidad de los errores

[8] Cuando existe autocorrelación, los errores tienden a ser seguidos por errores del mismo signo. Esta inercia es a la que se refiere la autocorrelación.

[9] Lo importante para que se cumplan las buenas propiedades del estimador de MCO es que se cumpla lo siguiente: $cov(X, u) = 0$, es decir, que no exista correlación entre X y el término de error.

Es conveniente asumir que los errores se distribuyen independiente e idénticamente (iid) con media cero y varianza constante. Es decir:

$$u_i \sim N(0, \sigma^2) \tag{13}$$

Este supuesto no es necesario para llevar a cabo la estimación de MCO, pero sí se requiere para hacer inferencia.

Supuesto 7: Especificación correcta del modelo
Un último supuesto se refiere al hecho que el modelo de regresión debe estar especificado de forma correcta. Ahora bien, este supuesto podría violarse básicamente por tres razones: (i) inclusión de regresores irrelevantes o exclusión de regresores relevantes, (ii) no linealidad del modelo, o (iii) parámetros inestables.

Una vez establecidos los supuestos bajo los cuales se construye el modelo clásico de regresión lineal, es posible pasar a considerar los criterios para la elección de un buen estimador.

2.2. Criterios para la elección de un estimador

A continuación se describen los criterios que se utilizan para elegir un buen estimador. Además, se mencionará brevemente cómo se comporta el estimador de mínimos cuadrados ordinarios (MCO) en cada uno de estos criterios.

Costo computacional. Un aspecto no menor al momento de decidir utilizar un determinado estimador se refiere a la factibilidad de poder implementarlo. En efecto, pudiera existir un estimador con excelentes propiedades estadísticas, sin embargo, si no se contara con los medios necesarios para poder implementarlo, como por ejemplo, un buen computador, no sería de mucha utilidad. Afortunadamente, hoy en día existen poderosos equipos, además de sofisticados softwares (Eviews, Gauss, RATS, SAS, SPSS, STATA, SHAZAM), que permiten la implementación eficiente, tanto en términos de tiempo como de costo, de los distintos tipos de estimadores. El estimador de mínimos cuadrados ordinarios (MCO) es muy poco exigente en términos computacionales.

Mínimos cuadrados. Otra característica deseable en un estimador es que minimice, de una manera u otra, la desviación de la línea estimada de los puntos muestrales observados. El estimador MCO minimiza precisamente la suma al cuadrado de los desvíos de los datos respecto de la función estimada.

Bondad de ajuste. Un buen estimador debiera ser capaz de proveer un buen grado de ajuste, por lo tanto, esta característica es muy deseable. El estimador MCO al minimizar la suma al cuadrado de los errores estimados, maximiza el valor para el R^2.

Insesgamiento. Si bien es imposible que un estimador proporcione el verdadero valor del parámetro poblacional (β), sí es muy importante que en promedio la estimación arroje un valor cercano al verdadero. En otras palabras, si se contara con muchas muestras distintas, entonces, se esperaría que en promedio los coeficientes estimados coincidieran con el verdadero valor del parámetro poblacional. Esta propiedad se denomina insesgamiento, y con todos los supuestos que se han hecho hasta ahora, se asegura el insesgamiento del estimador MCO.

Eficiencia. Sería apropiado también que la estimación sea lo más precisa posible, o en otras palabras, que tenga la menor varianza asociada. Dado que todo estimador tiene un intervalo de confianza asociado, lo ideal sería que dicho intervalo fuera lo más angosto posible. El estimador MCO, con los supuestos realizados hasta ahora, constituye el estimador más eficiente de entre todos los estimadores lineales e insesgados (teorema de Gauss-Markov).

Error cuadrático medio (ECM). Respecto de las propiedades de insesgamiento y eficiencia de un estimador, cabe señalar que mientras que un investigador pudiera preferir un estimador insesgado pero de mayor varianza, otro preferiría el de menor varianza, sin importar si es sesgado o no. Ahora bien, en este contexto, pudiera ser conveniente tomar en cuenta ambos criterios, es decir, eficiencia e insesgamiento. El ECM permite hacer esto, ya que en su cálculo considera tanto el sesgo del estimador como su varianza. El estimador MCO presenta el mejor ECM por las razones dadas anteriormente.

Propiedades asintóticas. Es probable que cierto estimador no cumpla con algunas propiedades deseables de muestras pequeñas, como por ejemplo, insesgamiento. Sin embargo, su utilización podría justificarse sobre la base de su comportamiento en muestras grandes, en donde se podrían satisfacer algunas propiedades deseables, como por ejemplo, consistencia y eficiencia asintótica. El estimador MCO, bajo los supuestos realizados, es consistente y asintóticamente eficiente.

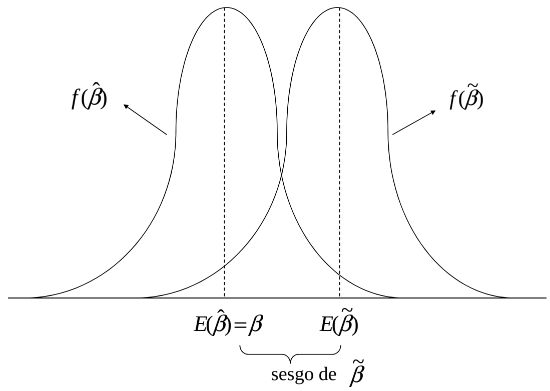
Máxima verosimilitud (MV). Finalmente, otro criterio que pudiera ser considerado al momento de seleccionar un estimador es el de la máxima verosimilitud, que consiste en ver qué tan bien permite el estimador seleccionado caracterizar la muestra con la que se está trabajando. Cabe señalar que cuando se asume normalidad en el término de error, y además, si el modelo es lineal, entonces, el estimador de MV coincide con el estimador MCO.

2.3. Propiedades de un estimador

Un buen estimador es aquel que exhibe las siguientes propiedades.

Insesgamiento. Se dice que $\hat{\beta}$ es un estimador insesgado del parámetro poblacional β si $E(\hat{\beta}) = \beta$. Esta propiedad establece que, utilizando varias muestras de datos distintas, en promedio el estimador corresponde al verdadero valor del parámetro poblacional.

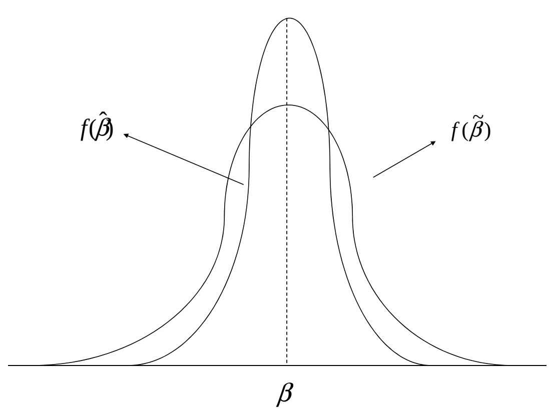
Figura 3: Utilizando la distribución muestral para ilustrar el sesgo



En otras palabras, un estimador se dice ser insesgado si el valor promedio de éste ($\hat{\beta}$) en muestras repetidas es igual al parámetro poblacional (β). En la figura 3 se aprecia que el estimador $\hat{\beta}$ tiene un valor esperado que es igual al parámetro poblacional β, distribuyéndose en torno a éste. En cambio, el estimador $\tilde{\beta}$ tiene un valor esperado mayor, y por ende, es sesgado (hacia arriba). Note que en este caso particular se ha asumido que ambos estimadores tienen la misma varianza.

Eficiencia o mínima varianza Se dice que $\hat{\beta}$ es un estimador eficiente si es que no existe ningún otro estimador insesgado con menor varianza. Así, de entre un conjunto de estimadores insesgados se escogerá aquel que tenga la menor varianza, y a éste se le conoce como el mejor estimador, o el estimador eficiente.[10]

Figura 4: Eficiencia relativa



La figura 4 muestra como el estimador $\hat{\beta}$ es más eficiente en términos relativos que el estimador $\tilde{\beta}$. Al estimador que es lineal e insesgado, y que además tiene la

[10]En la práctica es posible que no exista un estimador insesgado para un determinado parámetro, por lo que en esos casos se podría escoger aquel de menor ECM.

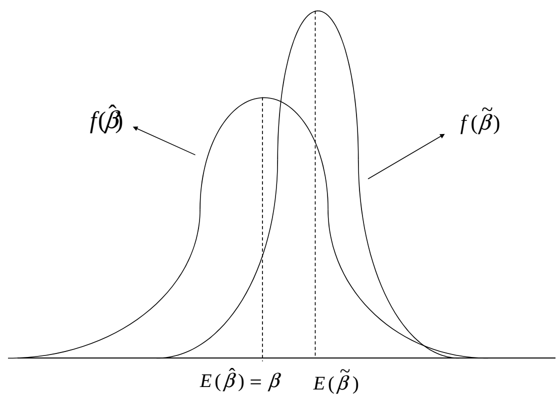
mínima varianza de entre todos los estimadores lineales e insesgados, se le denomina el mejor estimador lineal e insesgado (MELI).[11]

Error cuadrático medio. El error cuadrático medio (ECM) combina los conceptos de eficiencia e insesgamiento, lo que resulta muy útil al momento de escoger entre estimadores sesgados, y se define de la siguiente forma (en el caso de un escalar):

$$ECM(\hat{\beta}) = E[(\hat{\beta} - \beta)^2] = var(\hat{\beta}) + [sesgo(\hat{\beta})]^2$$

En palabras, es posible definir al ECM como una medida de dispersión en torno al verdadero valor del parámetro.

Figura 5: Error cuadrático medio



Tal como se puede aprecir en la figura 5, si bien el estimador $\hat{\beta}$ es insesgado, tiene asociado una mayor varianza. En cambio, el estimador $\tilde{\beta}$ posee una menor varianza, pero es sesgado. La elección sobre qué estimador es preferido se puede obtener a través de la construcción de un estadístico que incorpore tanto el sesgo como la varianza del estimador, asignándole una cierta ponderación a cada uno. Si los ponderadores son iguales, entonces, se obtiene el ECM. En la práctica el ECM no es muy utilizado a menos que el criterio del mejor estimador insesgado no sea capaz de producir estimadores con varianzas pequeñas.

Consistencia. En la práctica ocurre que los estimadores no son insesgados, sobretodo cuando se trabaja con muestras pequeñas. ¿Significa esto que deba descartarse dicho estimador? No. Si esto ocurre, los econometristas pueden justificar el uso de un estimador sobre la base de sus propiedades asintóticas, es decir, en función de su comportamiento en un contexto (teórico) en donde el tamaño muestral tiende a infinito ($n \to \infty$). Si la distribución asintótica del estimador comienza a concentrarse en un valor particular (β) a medida que la muestra tiende a infinito, β se dice ser la probabilidad límite del estimador; surge así el concepto de consistencia. El estimador $\hat{\beta}$ es consistente si converge (en el límite) al verdadero valor del parámetro

[11] BLUE por sus siglas en inglés.

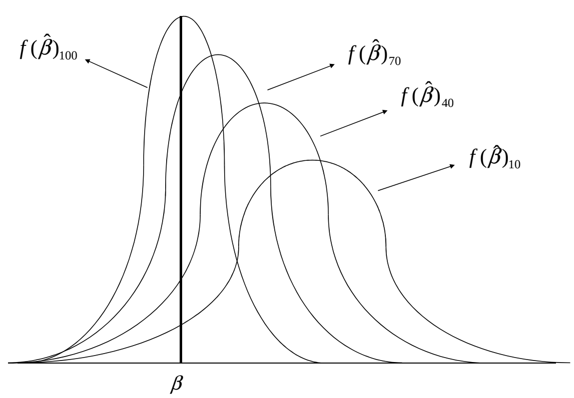
poblacional (β). Hay distintas formas de demostrar convergencia, en particular, la convergencia en probabilidad de un estimador se define de la siguiente manera:

$$\lim_{n\to\infty} Pr(|\hat{\beta}-\beta|<\epsilon)=1 \quad \forall\epsilon>0$$

La convergencia en probabilidad, denominada también *plim*, es el equivalente al insesgamiento, pero en términos asintóticos. Si el estimador es consistente, entonces, a medida que se incrementa el tamaño de la muestra, la distribución muestral del estimador tenderá a una línea vertical.[12]

Considere el estimador $\hat{\beta}$, calculado a partir muestras de distintos tamaños (n_{10}= 10 datos, n_{40}= 40 datos, y así sucesivamente). Si la distribución del estimador $\hat{\beta}$ se asimila cada vez más a alguna distribución específica a medida que crece el tamaño de la muestra (como una distribución normal por ejemplo), entonces, a esta distribución se le denomina distribución asintótica de $\hat{\beta}$. La figura 6 muestra conceptualmente lo expuesto.

Figura 6: Distribución muestral y tamaño de la muestra



La varianza de la distribución asintótica de $\hat{\beta}$ se denomina varianza asintótica. Si $\hat{\beta}$ es consistente y su varianza asintótica es más pequeña que la de otros estimadores consistentes, entonces, se dice que $\hat{\beta}$ es asintóticamente eficiente. La varianza de un estimador asintóticamente eficente converge a cero más rápido que la varianza de cualquier otro estimador consistente. Tal como lo muestra la figura 6, si $\hat{\beta}$ fuese consistente, entonces, a medida que la muestra crece a infinito, la distribución muestral llegaría a ser simplemente una línea vertical, situada exactamente en el valor de β.

Ahora que ya se han analizado las propiedades que debe cumplir todo buen estimador, la siguiente sección presenta el estimador de mínimos cuadrados ordinarios, un estimador ampliamente utilizado en la práctica por las buenas propiedades que exhibe.

[12] Por lo tanto, es posible concluir que un estimador insesgado es consistente, pero un estimador consistente no necesariamente es insesgado.

3. El estimador de mínimos cuadrados ordinarios

El estimador de mínimos cuadrados ordinarios (MCO) es el método tradicional de estimación. Así, MCO permite estimar los parámetros del siguiente modelo:

$$y = X\beta + u \tag{14}$$

Luego:

$$\hat{y} = X\hat{\beta} \tag{15}$$

De la ecuación (14) es posible notar que el error (teórico) viene dado por la siguiente expresión:

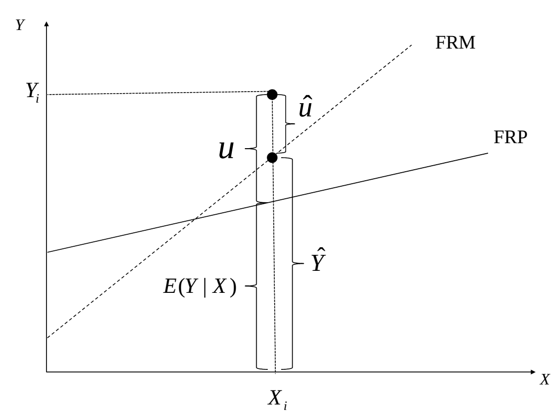
$$u = Y - X\beta \tag{16}$$

Por lo tanto, con la estimación que se realiza del vector de parámetros ($\hat{\beta}$) es posible notar que el vector de errores de la estimación será:

$$\hat{u} = Y - X\hat{\beta}$$

El problema consiste en encontrar el vector de parámetros ($\hat{\beta}$) que minimice la suma de los errores al cuadrado. Dado que algunos errores serán positivos (la línea ajustada pasa por debajo de la línea verdadera) y otros serán negativos (la línea ajustada pasará por encima de la línea verdadera), entonces se toma la suma al cuadrado. La figura 7 permite visualizar el problema que se ha planteado pero en donde existen solo dos variables, X e y.

Figura 7: El método de mínimos cuadrados ordinarios



Así, intuitivamente, lo que hace MCO es encontrar la recta que mejor ajusta la nube de puntos observada. En este contexto, FRP denota la función de regresión poblacional, que es la relación teórica no observada entre X e y, que en este caso se ha asumido es linal (FRP: $E(y|X) = \beta_1 + \beta_2 X$). Esta relación teórica se estima por medio de la función de regresión muestral (FRM) que corresponde al modelo estimado:

$$\hat{E}(y|X) = \hat{y} = \hat{\beta}_1 + \hat{\beta}_2 X$$

Es interesante notar que ninguna de las rectas (FRM,FRP) pasa por el punto verdadero, el dato observado (X_i,y_i), ya que ambas especificaciones tienen un componente estocástico.[13] De esta manera, para el caso de la FRP se tiene que:

$$y_i = FRP + u_i = E(y|X) + u_i$$

y para la FRM se tiene que:

$$y_i = FRM + u_i = \hat{y} + \hat{u}_i$$

De esta manera, el error de MCO ($\hat{u}$) es el estimador de u (el shock estocástico no observable de la FRP), y $\hat{y} = X\hat{\beta}$ es el estimador de $E(y|X)$. En términos más generales, la función objetivo que se quiere optimizar se puede plantear en términos matriciales, para lo cual se debe notar lo siguiente:

$$\sum_{i=1}^{n} u_i^2 = u'u \tag{17}$$

Por lo tanto, el problema consiste en minimizar la siguiente expresión que corresponde a la suma de los errores al cuadrado:

$$S(\beta) = u'u = (y - X\beta)'(y - X\beta) \tag{18}$$

Para encontrar un punto mínimo en la función objetivo anterior, se deriva con respecto al vector de parámetros (β) y se iguala a cero (note que esta expresión corresponde a un vector cuyos elementos son las derivadas con respecto a β_1,β_2,...,β_k):

$$\frac{\partial S(\beta)}{\partial \beta} = -2X'(y - X\hat{\beta}) = 0$$

Despejando:

$$X'y = X'X\hat{\beta} \tag{19}$$

A este conjunto de k ecuaciones se les denomina ecuaciones normales. Si la inversa de la matriz $X'X$ existe, lo cual se cumple pues X debe tener rango completo, la solución a este problema viene dada por la siguiente expresión:

$$\hat{\beta} = (X'X)^{-1}X'y \tag{20}$$

Por lo tanto, a $\hat{\beta}$ se le conoce como el estimador de mínimos cuadrados ordinarios. Por otro lado, note que la matriz de segundas derivadas viene dada por:

$$\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta'} = 2X'X \tag{21}$$

Luego, para que la solución a este sistema de ecuaciones corresponda a un mínimo, la matriz ($X'X$) debe ser definida positiva, es decir, sus valores propios deben ser mayores o iguales a cero, lo cual se cumple.[14]

[13] No obstante, más adelante se mostrará que la estimación por MCO siempre pasa por el punto medio de la muestra, es decir, pasa por el punto ($\bar{X}$,$\bar{y}$).

[14] Los valores propios de una matriz A se obtienen solucionando lo siguiente:

$$|A - \lambda I| = 0$$

3.1. Caso particular: MCO en el modelo de regresión lineal de dos variables

Considere el siguiente modelo bivariado, en donde la variable dependiente (Y) se explica solo por una variable independiente (X):

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Luego, los errores de MCO vienen dados por:

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i$$

Por lo tanto, la suma de los errores al cuadrado corresponde a:

$$S(\beta) = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$$

Las condiciones de primer orden (CPO) son las siguientes (son dos pues se debe derivar con respecto a $\hat{\beta}_1$ y $\hat{\beta}_2$):

$$\frac{\partial S(\beta)}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-1) = 0 \Rightarrow \sum_{i=1}^{n} \hat{u}_i = 0$$

y:

$$\frac{\partial S(\beta)}{\partial \hat{\beta}_2} = \sum_{i=1}^{n} 2(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(-X_i) = 0 \Rightarrow \sum_{i=1}^{n} X_i \hat{u}_i = 0$$

Por lo tanto, las ecuaciones normales vienen dadas por:

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_1 + (\sum_{i=1}^{n} X_i)\hat{\beta}_2$$

$$\sum_{i=1}^{n} X_i Y_i = (\sum_{i=1}^{n} X_i)\hat{\beta}_1 + (\sum_{i=1}^{n} X_i^2)\hat{\beta}_2$$

A modo de ejemplo, considere la siguiente matriz:

$$A = \begin{pmatrix} 1.3 & -0.1 \\ 0.8 & 0.4 \end{pmatrix}$$

Por lo tanto, los valores propios se obtienen de solucionar:

$$\left| \begin{pmatrix} 1.3 & -0.1 \\ 0.8 & 0.4 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = \left| \begin{pmatrix} 1.3 - \lambda & -0.1 \\ 0.8 & 0.4 - \lambda \end{pmatrix} \right| = 0$$

Luego:

$$\lambda^2 - 1.7\lambda + 0.6 = 0$$

cuyas soluciones son λ_1=1.2 y λ_2=0.5, que corresponden a los valores propios de A.

Note que este sistema consta de dos ecuaciones y dos incógnitas, por lo tanto, resulta sencillo encontrar la solución. Para ello se divide la primera ecuación por n, por lo que se llega a:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}$$

De la expresión anterior, es posible notar que la estimación de MCO pasa por los puntos medios muestrales de X e Y. Luego, el estimador MCO para el intercepto de la ecuación es:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}$$

Luego, reemplazando $\hat{\beta}_1$ en la segunda ecuación normal:

$$\begin{aligned}
\sum X_i Y_i &= (\sum X_i)(\bar{Y} - \hat{\beta}_2 \bar{X}) + (\sum X_i^2)\hat{\beta}_2 \\
\sum X_i Y_i &= \sum X_i \bar{Y} - \sum X_i \hat{\beta}_2 \bar{X} + (\sum X_i^2)\hat{\beta}_2 \\
\sum X_i Y_i &= n\bar{X}\bar{Y} + \hat{\beta}_2(\sum X_i^2 - n\bar{X}^2)
\end{aligned}$$

Finalmente, el estimador MCO para β_2 viene dado por:

$$\hat{\beta}_2 = \frac{(\sum X_i Y_i) - n\bar{X}\bar{Y}}{(\sum X_i^2) - n\bar{X}^2}$$

Alternativamente, se puede mostrar lo siguiente. Recuerde que de la segunda ecuación normal se tiene:

$$\sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)(X_i) = 0$$

Luego:

$$\sum X_i (Y_i - (\bar{Y} - \hat{\beta}_2 \bar{X}) - \hat{\beta}_2 X_i) = 0$$

$$\sum X_i (Y_i - \bar{Y} + \hat{\beta}_2 \bar{X} - \hat{\beta}_2 X_i) = 0$$

$$\sum X_i (Y_i - \bar{Y} - \hat{\beta}_2 (X_i - \bar{X})) = 0$$

$$\sum X_i (Y_i - \bar{Y} - \hat{\beta}_2 \sum X_i (X_i - \bar{X})) = 0$$

Haciendo uso de la siguiente propiedad:

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = \sum X_i (Y_i - \bar{Y}) \tag{22}$$

Se tiene que:

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = \hat{\beta}_2 \sum (X_i - \bar{X})(X_i - \bar{X}) = 0$$

Finalmente:

$$\hat{\beta}_2 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2} \tag{23}$$

donde la variable en minúscula indica que se encuentra en desvío con respecto a su media. Es directo verificar que las condiciones de segundo orden para un mínimo se cumplen en este problema. En efecto, la matriz de segundas derivadas viene dada por:

$$H = \begin{pmatrix} \frac{\partial^2 \sum \hat{u}^2}{\partial \hat{\beta}_1^2} & \frac{\partial^2 \sum \hat{u}^2}{\partial \hat{\beta}_2 \partial \hat{\beta}_1} \\ \frac{\partial^2 \sum \hat{u}^2}{\partial \hat{\beta}_1 \partial \hat{\beta}_2} & \frac{\partial^2 \sum \hat{u}^2}{\partial \hat{\beta}_2^2} \end{pmatrix} = \begin{pmatrix} 2n & 2n\bar{X} \\ 2n\bar{X} & 2\sum X_i^2 \end{pmatrix}$$

Para que la matriz sea semidefinida positiva los valores propios de H deben ser positivos. Además, dado que el determinante de una matriz corresponde a la pitatoria de los valores propios, es decir:

$$|H| = \Pi_{j=1}^{J} \lambda_j$$

entonces, el determinante H debe ser positivo. Luego, en este caso:

$$|H| = 2n(2\sum X_i^2) - (2n\bar{X})^2 = 4n(\sum X_i^2 - n\bar{X}^2) = 4n\sum (X_i - \bar{X})^2 \geq 0$$

La tabla 1 presenta los resultados de una ecuación de Mincer (ecuación de salarios) utilizando una muestra de los datos proveniente de la encuesta de Caracterización Socioeconómica Nacional (CASEN) del año 2009 utilizando el software EVIEWS.[15] La ecuación de Mincer es ampliamente utilizada en la econometría empírica para analizar los determinantes de los salarios de los individuos. El modelo a estimar es el siguiente:

$$ln(Y_i) = \beta_1 + \beta_2 ESC_i + u_i$$

La variable dependiente ($ln(Y)$) corresponde al logaritmo del salario por hora, y la variable independiente (ESC) son los años de escolaridad, que constituye una variable proxy del capital humano de la persona. El shock (u) representa todos aquellos factores no observables por parte del econometrista pero que afectan el salario del individuo (esfuerzo, habilidad, etc.). Dado que la variable dependiente está medida en logaritmos y la variable independiente está en nivel (número de años), el coeficiente estimado corresponde a una semi-elasticidad. De esta manera, la estimación revela que por cada año adicional de escolaridad el salario por hora de las personas crece en promedio un 8.02 %.

Tabla 1: Ecuación de Mincer
Dependent Variable:LWPH
Method: Least Squares
Sample: 12900
Included observations: 2900

[15] La encuesta CASEN es realizada periódicamente por el Ministerio de Planificación (MIDEPLAN) de Chile. Esta encuesta se ha llevado a cabo a partir del año 1985, siendo el año 2011 su más reciente versión. Su principal objetivo es realizar un diagnóstico de la situación socioeconómica de los hogares y de la población del país, para lo cual recopila para cada uno de los integrantes de los hogares seleccionados información relativa a empleo, salud, vivienda, e ingresos.

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	6.258460	0.031904	196.1681	0.0000
ESC	0.080275	0.002825	28.41752	0.0000
R-squared	0.217931	Mean dependent var		7.112859
Adjusted R-squared	0.217661	S.D dependent var		0.649708
S.E of regression	0.574666	Akaike info criterion		1.730634
Sum squared resid	957.0385	Schwarz criterion		1.734753
Log likelihood	-2507.419	Hannan-Quinn criter.		1.732118
F-statistic	807.5552	Durbin-Watsons stat		2.069329
Prob (F-statistic)	0.000000			

3.2. Aspectos algebraicos

Volviendo al caso general, a continuación se consideran algunos aspectos algebraicos importantes de tener en cuenta. Primero, recuerde que las ecuaciones normales vienen dadas por:

$$X'X\hat{\beta} = X'y$$

Por lo tanto:

$$\begin{aligned} X'X\hat{\beta} - X'y &= 0 \\ -X'(y - X\hat{\beta}) &= 0 \\ -X'\hat{u} &= 0 \\ X'\hat{u} &= 0 \end{aligned}$$

Esto quiere decir que para cada una de las columnas de X, es decir, para cada uno de los regresores incluidos, se cumple que $X_k'\hat{u} = 0$. En otras palabras, X es ortogonal (independiente) al error estimado por MCO ($\hat{u}$). Ahora bien, note que si la primera columna de X es una columna de unos[16], entonces, hay tres implicancias de la expresión anterior:

1. La suma de los errores de mínimos cuadrados ordinarios es cero. Esto sigue del hecho que la primera columna de la matriz X es de unos, pues el modelo incluye una constante. Por lo tanto, se tiene que: $X_1'\hat{u} = i'\hat{u} = \sum_i^n \hat{u}_i = 0$, donde i' es un vector fila de unos.

2. La regresión pasa por el punto medio de los datos, ya que al dividir la primera ecuación normal por n se obtiene lo siguiente: $\bar{y} = \bar{X}'\hat{\beta}$

3. El promedio de los valores ajustados de la regresión iguala el promedio de los valores de Y

Se sabe que los errores MCO vienen dados por la siguiente expresión:

$$\hat{u} = y - \hat{y} = y - X\hat{\beta} \tag{24}$$

[16]Esto se debe a que el modelo incluye una constante.

Luego, reemplazando $\hat{\beta}$:

$$\hat{u} = y - X(X'X)^{-1}X'y = (I - X(X'X)^{-1}X')y$$

Por lo tanto:

$$\hat{u} = MY \tag{25}$$

La matriz M, de dimensión nxn, es fundamental en el análisis de regresión.[17] Así, la matriz M al premultiplicar el vector Y entrega el vector de residuos provenientes de la regresión de Y contra X.[18] Por otro lado, se tiene que:

$$\hat{y} = y - \hat{u} = (I - M)y = Py$$

La matriz P (simétrica e idempotente) se denomina matriz de proyección. Esta matriz al premultiplicar el vector y entrega los valores ajustados de la regresión de MCO de y en contra de X, es decir, entrega $\hat{y}$.[19]

Finalmente se tienen las siguientes expresiones equivalentes para la suma de los errores al cuadrado:

$$\begin{aligned} \hat{u}'\hat{u} &= y'M'My = y'My = y'\hat{u} = \hat{u}'y \\ \hat{u}'\hat{u} &= y'y - \hat{\beta}'X'X\hat{\beta} \\ \hat{u}'\hat{u} &= y'y - \hat{\beta}'X'y = y'y - y'X\hat{\beta} \end{aligned}$$

3.3. Propiedades del estimador de MCO: El Teorema de Gauss-Markov

En la práctica, el econometrista debe escoger entre distintos métodos alternativos de estimación. Es por ello que la elección de uno u otro método debiese depender de las propiedades estadísticas de los estimadores que se tienen como candidatos, tales como el insesgamiento, la eficiencia, la precisión, e incluso el costo computacional. Lo interesante es que el estimador de mínimos cuadrados ordinarios (MCO) presenta excelentes propiedades, sin siquiera tener que especificar una distribución particular para el término de error (u) del modelo. En efecto, recuerde que:

$$\hat{\beta} = (X'X)^{-1}X'y$$

[17] Es posible mostrar que M es simétrica ($M = M'$) e idempotente ($M = M^2$).

[18] Por lo tanto:

$$MX = 0$$

[19] Se cumple que:

$$PX = X$$

y:

$$PM = MP = 0$$

Además, note que:

$$\hat{y} = Py + My$$

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u)$$
$$\hat{\beta} = \beta + (X'X)^{-1}X'u \quad (26)$$

Por lo tanto, si X es no estocástico, o bien, si $E(u|X) = 0$, se cumple lo siguiente:[20]

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(u) = \beta \quad (27)$$

De esta manera, sin importar la distribución del término de error (u), $\hat{\beta}$ es un estimador insesgado de β. Es decir, en promedio (si se tuvieran muchas muestras repetidas para X e y, y se calculara el promedio de $\hat{\beta}$ para todas ellas) la estimación MCO ($\hat{\beta}$) corresponde al verdadero parámetro poblacional (β). Por otro lado, la matriz de varianzas y covarianzas del estimador MCO (note que es un vector de parámetros) es la siguiente:

$$E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))']$$

Como el estimador MCO es insesgado, se llega a:

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

Por otro lado, de (26) se tiene que:

$$\hat{\beta} - \beta = (X'X)^{-1}X'u \quad (28)$$

entonces:

$$var(\hat{\beta}) = E[(X'X)^{-1}X'uu'X(X'X)^{-1}]$$
$$var(\hat{\beta}) = (X'X)^{-1}X'E(uu')X(X'X)^{-1}$$
$$var(\hat{\beta}) = (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1}$$

Por lo tanto, la matriz de varianzas y covarianzas queda de la siguiente forma:

$$E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \sigma^2(X'X)^{-1} \quad (29)$$

A continuación se obtendrá un resultado muy importante para la clase de estimadores lineales e insesgados de β. Sea $b_0 = CY$ cualquier otro estimador lineal e insesgado de β, donde C es una matriz de kxn. Si b_0 es insesgado, entonces, tiene que cumplirse que:

$$E(CY) = E((CX\beta + Cu) = \beta$$

Es decir, debe darse que $CX = I$. Por lo tanto, como se puede apreciar, existen muchos candidatos potenciales para estimar el vector de parámetros poblacionales. La matriz de varianzas y covarianzas de b_0 se puede encontrar reemplazando $(X'X)^{-1}X'$ por C en la ecuación (26), por lo tanto:

$$var(b_0) = \sigma^2 CC'$$

[20] Con este último supuesto se demuestra que insesgamiento condicional implica insesgamiento incondicional, utilizando la ley de expectativas iteradas.

Sea $D = C - (X'X)^{-1}X'$ tal que $DY = b_0 - \hat{\beta}$. De esta manera, la varianza de b_0 puede escribirse de la siguiente manera:

$$var(b_0) = \sigma^2[(D + (X'X)^{-1}X')(D + (X'X)^{-1}X')']$$
$$var(b_0) = \sigma^2[(D + (X'X)^{-1}X')(D' + X(X'X)^{-1})]$$
$$var(b_0) = \sigma^2(DD' + DX(X'X)^{-1} + (X'X)^{-1}X'D' + (X'X)^{-1}(X'X)(X'X)^{-1})$$
$$var(b_0) = \sigma^2(DD' + DX(X'X)^{-1} + (X'X)^{-1}X'D' + (X'X)^{-1})$$

ya que $C = D+(X'X)^{-1}X'$. Por otro lado, se sabe que $CX = DX+(X'X)^{-1}(X'X) = DX + I$, pero además, se asumió que $CX = I$, por lo que necesariamente tiene que cumplirse que $DX = 0$. Por lo tanto:

$$var(b_0) = \sigma^2(X'X)^{-1} + \sigma^2 DD'$$

es decir:

$$var(b_0) = var(\hat{\beta}) + \sigma^2 DD' \tag{30}$$

Así, la matriz de varianzas y covarianzas del estimador b_0 equivale a la de $\hat{\beta}_{MCO}$ más una matriz que es definida no negativa, lo cual da origen a una importante conclusión respecto del estimador de MCO, que se resume en el siguiente teorema.

Teorema de Gauss-Markov: en el modelo clásico de regresión lineal el estimador MCO $\hat{\beta}$ es el estimador lineal insesgado de mínima varianza de β (MELI).[21]

En resumen, el estimador de MCO es una poderosa técnica de estimación pues proporciona estimadores que cumplen con propiedades estadísticas deseables, en la medida que se cumplan ciertos supuestos. En casos más generales que el considerado hasta ahora, en donde algunos de los supuestos acá mencionados no se cumplan, el campo es mucho más amplio, y hay espacio para que otros estimadores compitan favorablemente contra MCO.

3.3.1. Distribución del estimador MCO

Hasta el momento no se ha hecho uso del supuesto de normalidad del término de error (u), ya que la derivación del estimador MCO prescinde completamente de este elemento. Sin embargo, este supuesto es fundamental para realizar inferencia estadística (test de hipótesis). Dado que $\hat{\beta}$ depende de Y, el que a su vez está en función de u, entonces, $\hat{\beta}$ es estocástico, y por ende, tiene asociada la distribución de u. Por lo tanto, la distribución de $\hat{\beta}$ viene dada por:

$$\hat{\beta} \sim N(\beta, \sigma^2(X'X)^{-1}) \tag{31}$$

y cada elemento $\hat{\beta}_j$ del vector $\hat{\beta}$ tiene la siguiente distribución:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2(X'X)^{-1}_{jj}) \tag{32}$$

[21] El estimador MCO tiene la mínima varianza de entre todos los estimadores lineales e insesgados. Sin embargo, note que es posible que existan estimadores que aún siendo sesgados presenten un menor ECM.

donde $(X'X)^{-1}_{jj}$ corresponde al j-ésimo elemento de la diagonal de dicha matriz (que es de kxk). Algo que es importante de notar en este punto es que la distribución de $\hat{\beta}$ en muestras finitas es consecuencia directa del supuesto específico de que los errores tienen una distribución normal. Ahora bien, si el término de error no se distribuye normal aún así es posible asumir normalidad en la distribución de $\hat{\beta}$ (siempre y cuando la muestra tenga un tamaño razonable) apelando al Teorema Central del Límite.

En el contexto del modelo de dos variables se sabe que:[22]

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i(\beta_2 x_i + u_i)}{\sum_{i=1}^n x_i^2}$$

Luego:

$$\hat{\beta}_2 = \beta_2 + \frac{\sum_{i=1}^n x_i u_i}{\sum_{i=1}^n x_i^2} \tag{33}$$

Aplicando el operador esperanza a la expresión anterior, se obtiene lo siguiente:

$$E(\hat{\beta}_2) = E(\beta_2) + E\left(\frac{\sum_{i=1}^n x_i u_i}{\sum_{i=1}^n x_i^2}\right) = \beta_2 + \frac{\sum x_i E(u_i)}{\sum x_i^2}$$

Finalmente:

$$E(\hat{\beta}_2) = \beta_2$$

es decir, $\hat{\beta}_2$ es un estimador insesgado de β_2. Por otro lado:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$$

$$\hat{\beta}_1 = \beta_1 + \beta_2\bar{X} + \bar{u} - \hat{\beta}_2\bar{X}$$

$$\hat{\beta}_1 = \beta_1 - (\hat{\beta}_2 - \beta_2)\bar{X} + \bar{u} \tag{34}$$

Aplicando valor esperado:

$$E(\hat{\beta}_1) = E(\beta_1) - E(\hat{\beta}_2 - \beta_2)\bar{X} + E(\bar{u}) = \beta_1 - (E(\hat{\beta}_2) - \beta_2)\bar{X} + E(\bar{u})$$

Finalmente:

$$E(\hat{\beta}_1) = \beta_1$$

es decir, $\hat{\beta}_1$ es un estimador insesgado de β_1. A continuación se consideran los segundos momentos de la distribución del estimador MCO, es decir, los que tienen que ver con la varianza. La varianza de $\hat{\beta}_2$ se calcula como sigue.

$$var(\hat{\beta}_2) = E(\hat{\beta}_2 - E(\hat{\beta}_2))^2 = E(\hat{\beta}_2 - \beta_2)^2$$

De (33), se reemplaza:

$$var(\hat{\beta}_2) = E\left(\frac{\sum_{i=1}^n x_i u_i}{\sum_{i=1}^n x_i^2}\right)^2$$

[22]Recuerde que la variable en minúscula denota que se encuentra en desvíos con respecto a la media.

Por lo tanto:

$$var(\hat{\beta}_2) = \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} E(\sum_{i=1}^{n} x_i u_i)^2 = \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} E(x_1u_1 + x_2u_2 + \cdots + x_nu_n)^2$$

Desarrollando la expresión:

$$var(\hat{\beta}_2) = \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} E(x_1^2u_1^2 + x_2^2u_2^2 + \cdots + x_n^2u_n^2 + 2x_ix_ju_iu_j) \quad (i < j)$$

Luego:

$$var(\hat{\beta}_2) = \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} (x_1^2E(u_1^2) + x_2^2E(u_2^2) + \cdots + x_n^2E(u_n^2) + 2x_ix_jE(u_iu_j))$$

Dado que los errores son homocedásticos ($E(u_i^2) = \sigma^2$) y no están autocorrelacionados ($E(u_iu_j) = 0$):

$$var(\hat{\beta}_2) = \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} (\sigma^2x_1^2 + \sigma^2x_2^2 + \cdots + \sigma^2x_n^2)$$

Por lo tanto:

$$var(\hat{\beta}_2) = \frac{1}{(\sum_{i=1}^{n} x_i^2)^2} \sigma^2 \sum x_i^2$$

Finalmente:

$$var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \tag{35}$$

Para la varianza de $\hat{\beta}_1$, note que de (34) se tiene que:

$$\hat{\beta}_1 - \beta_1 = (\hat{\beta}_2 - \beta_2)\bar{X} + \bar{u}$$

Luego:

$$E(\hat{\beta}_1 - \beta_1)^2 = E[(\hat{\beta}_2 - \beta_2)\bar{X} + \bar{u}]^2$$

Desarrollando los términos:

$$E(\hat{\beta}_1 - \beta_1)^2 = E(\hat{\beta}_2 - \beta_2)^2\bar{X}^2 + E(\bar{u})^2$$

$$E(\hat{\beta}_1 - \beta_1)^2 = \bar{X}^2\frac{\sigma^2}{\sum x_i^2} + \frac{n\sigma^2}{n^2}$$

Finalmente:

$$var(\hat{\beta}_1) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} \right)$$

Finalmente, la covarianza entre $\hat{\beta}_1$ y $\hat{\beta}_2$ viene dada por:

$$\begin{aligned} cov(\hat{\beta}_1, \hat{\beta}_2) &= E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2)] \\ cov(\hat{\beta}_1, \hat{\beta}_2) &= E[(\bar{u} - (\hat{\beta}_2 - \beta_2)\bar{X})(\hat{\beta}_2 - \beta_2)] \\ cov(\hat{\beta}_1, \hat{\beta}_2) &= E[(\hat{\beta}_2 - \beta_2)] - \bar{X}E[(\hat{\beta}_2 - \beta_2)^2] \\ cov(\hat{\beta}_1, \hat{\beta}_2) &= -\frac{\sigma^2\bar{X}}{\sum_{i=1}^{n} x_i^2} \end{aligned}$$

3.3.2. El estimador de σ^2

Para poder testar hipótesis sobre el parámetro estimado, o bien, para construir intervalos de confianza, se requiere de una estimación de la matriz de varianzas y covarianzas del estimador MCO, que como se vio, viene dada por:

$$var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Lo anterior, exige contar con una estimación del parámetro poblacional σ^2, el cual es desconocido, y que corresponde a la varianza del término de error. Ahora bien, dado que:

$$var(u) = E(u - E(u))^2 = E(u^2) = \sigma^2$$

entonces, resulta natural estimar σ^2 utilizando la fórmula de la varianza de $\hat{u}$:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n}$$

Sin embargo, $\hat{\sigma}^2$ es un estimador sesgado, pero consistente, de la varianza del término de error. Lo anterior se debe al hecho de que los residuos de MCO sólo son estimadores imperfectos de los residuos poblacionales.[23] Con el objetivo de hallar un estimador insesgado para σ^2 recuerde lo siguiente:[24]

$$\hat{u} = MY = M(X\beta + u) = Mu$$

Por otro lado, es razonable considerar que un estimador de σ^2 debiera basarse en la suma de los residuos MCO al cuadrado. Por lo tanto:

$$\hat{u}'\hat{u} = (Mu)'Mu = u'M'Mu = u'Mu$$

ya que M es una matriz idempotente; aplicando el operador esperanza a la expresión anterior:

$$E(\hat{u}'\hat{u}) = E(u'Mu)$$

Aplicando el operador traza (tr), y recordando que la traza de una constante es

[23] En efecto:

$$\hat{u}_i = y_i - x_i'\hat{\beta} = x_i'\beta + u_i - x_i'\hat{\beta} = u_i - x_i'(\hat{\beta} - \beta)$$

Es decir, la estimación está distorsionada debido a la imposibilidad de observar β directamente.

[24] Se cumple que $MX = 0$.

la constante, se tiene lo siguiente:[25]

$$E(\hat{u}'\hat{u}) = E[tr(u'M)u] = E[tr(Muu')]$$

por propiedad $tr(AB) = tr(BA)$. Ahora bien, dado que M es una matriz que no contiene elementos estocásticos, entonces:

$$tr(ME[uu']) = tr(M\sigma^2 I) = \sigma^2 tr(M)$$

Por otro lado, la traza de M viene dada por:

$$\begin{aligned} tr(M) &= tr[I_n - X(X'X)^{-1}X'] \\ tr(M) &= tr(I_n) - tr[(X'X)^{-1}X'X] \\ tr(M) &= tr(I_n) - tr(I_k) \\ tr(M) &= n - k \end{aligned}$$

Por lo tanto:

$$E(\hat{u}'\hat{u}) = (n-k)\sigma^2$$

De esta manera, es posible apreciar que el estimador natural de los residuos se encuentra sesgado hacia abajo, aunque el sesgo desaparece a medida que aumenta el tamaño de la muestra (n), por lo tanto, es un estimador consistente. En efecto:

$$E(\hat{\sigma}^2) = E\left(\frac{\hat{u}'\hat{u}}{n}\right) = \frac{(n-k)}{n}\sigma^2 = \left(1 - \frac{k}{n}\right)\sigma^2$$

Así, cuando n tiende a infinito, $E(\hat{\sigma}^2)$ converge a σ^2. Luego, es directo notar que un estimador insesgado de σ^2 viene dado por:

$$\tilde{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n-k} \tag{36}$$

y se cumple que:

$$E(\tilde{\sigma}^2) = \sigma^2$$

[25] La traza de una matriz cuadrada de kxk es la suma de los elementos de su diagonal:

$$tr(A) = \sum_{i=1}^{k} a_{ii}$$

con A una matriz de kxk. Algunas propiedades importantes del operador traza son las siguientes:

- $tr(cA) = c(tr(A))$
- $tr(A') = tr(A)$
- $tr(A+B) = tr(A) + tr(B)$
- $tr(I_K) = K$
- $tr(AB) = tr(BA)$

Luego, con $\tilde{\sigma}^2$ es posible estimar la varianza del estimador MCO:[26]

$$var(\hat{\beta}) = \tilde{\sigma}^2(X'X)^{-1} \tag{37}$$

Así, la raíz cuadrada del j-ésimo elemento de la diagonal de esta matriz $[\tilde{\sigma}^2(X'X)^{-1}]_{jj}{}^{1/2}$ corresponde al error estándar del j-ésimo coeficiente estimado $(\hat{\beta}_j)$.

3.4. Bondad de ajuste de la regresión

A continuación se presenta una medida para la bondad de ajuste de la regresión, es decir, un estadístico que permite pronunciarse respecto de la calidad del ajuste de la estimación de MCO a los datos observados. En otras palabras, se establecerá un criterio que permita saber cuán bien ajusta el modelo a la muestra de datos. Intuitivamente, lo ideal sería que la estimación pasara muy cerca de los datos observados. El criterio adecuado consiste en analizar si la variación de la variable dependiente (Y) se explica o no por la variación de las variables explicativas (X).

A modo de ilustración considere el siguiente modelo:

$$Y = \beta_1 + \beta_2 X + \beta_3 Z + u$$

es decir, Y depende de dos variables explicativas, X y Z. La figura 8 utiliza el diagrama de Ballentine-Venn para graficar la relación existente entre estas tres variables. Los tamaños de los círculos representan las variaciones (totales) de las variables Y, X y Z.[27] En este caso, la variable Y es explicada por los regresores X y Z, lo que se grafica en el grado de traslape que existe entre Y, X y Z. Por su parte, los círculos X y Z tendrán algún grado de traslape reflejando la colinealidad existente entre ambas variables (zona de color rojo y naranjo). Note que si Y fuera regresionada sólo contra X, entonces, se utilizaría la información representada por la zona de color azul y rojo para estimar β_2. En cambio, si Y fuera regresionada sólo contra Z, entonces, se utilizaría la información representada por la zona de color verde y rojo para estimar β_3. Cuando Y es regresionada contra X y Z conjuntamente, el estimador MCO utiliza el área de color azul para estimar β_2, descartando la proveniente del área de color rojo. La información en el área de color azul corresponde a la variación de Y que se debe a la variación de X. Utilizando esta información se obtiene una estimación insesgada de β_2. La información en el área de color verde corresponde a la variación de Y que se debe a la variación de Z. Utilizando esta información se obtiene una estimación insesgada de β_3. La información del área de color rojo no se utiliza porque refleja la variación de Y que es determinada por la variación conjunta de X y Z.[28] Note que las áreas de color azul, rojo y verde, representan la

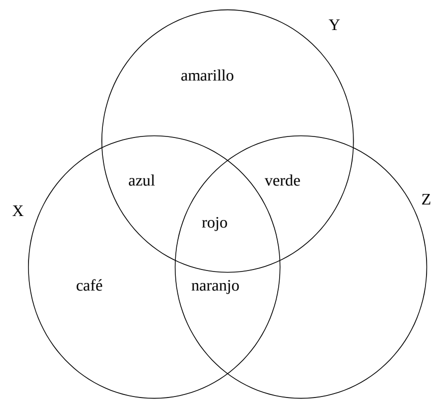
[26] Puede estimarse también con $\hat{\sigma}^2$. Cuando la muestra es grande no hay diferencias significativas entre utilizar uno u otro estimador de la varianza. Para muestras pequeñas puede ser más conveniente utilizar $\tilde{\sigma}^2$

[27] Por simplicidad se ha asumido que los tamaños de los círculos son iguales, es decir, que todas las variables tienen la misma variabilidad.

[28] Por lo tanto, se debe notar que al incorporar una mayor cantidad de regresores en la ecuación, se pierde precisión en las estimaciones, ya que deja de utilizarse información útil, en particular aquella

variación de Y que se explica por X y Z (en conjunto). Por lo tanto, una medida de la bondad de ajuste del modelo consistiría en calcular el cuociente entre la suma de las áreas de color rojo, verde y azul, y el círculo Y. En palabras, esta medida de ajuste indicaría el porcentaje de la variación de Y que puede atribuirse a X y Z. Finalmente, el área de color amarillo representa la variación de Y que se atribuye al término de error, y por lo tanto, la magnitud de dicha área representa la magnitud de la varianza del término de error.

Figura 8: El diagrama de Ballentine-Venn



Para construir un estadístico de la bondad de ajuste de la estimación, se necesita plantear el modelo en desvíos respecto a la media, para luego llevar a cabo un análisis de varianzas. En términos más formales, la variación total en Y, suma de cuadrados totales (STC), se define de la siguiente manera:

$$STC = \sum_{i=1}^{n}(Y_i - \bar{Y})^2 \tag{38}$$

Desarrollando la expresión anterior:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum[(Y_i - \hat{Y}_i) + (\hat{Y}_i - \bar{Y})]^2 = \sum[\hat{u}_i + (\hat{Y}_i - \bar{Y})]$$

Luego:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum[\hat{u}_i^2 + 2\hat{u}_i(\hat{Y}_i - \bar{Y}) + (\hat{Y}_i - \bar{Y})^2]$$

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum \hat{u}_i^2 + 2\sum \hat{u}_i(\hat{Y}_i - \bar{Y}) + \sum(\hat{Y}_i - \bar{Y})^2$$

que corresponde a la variación conjunta de los regresores y que también explica la variabilidad de Y. Con todo, si bien se pierde precisión al incorporar más variables en el análisis (aumentan las varianzas), las estimaciones siguen siendo insesgadas.

Finalmente:

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum \hat{u}_i^2 + \sum(\hat{Y}_i - \bar{Y})^2$$

Por lo tanto:

$$STC = SCE + SEC$$

donde SCE es la suma de los cuadrados de los errores, y SEC es la suma explicada de los cuadrados. Luego, el R^2 se define de manera natural como el porcentaje de la variación de Y que puede ser explicada por la variación de X:

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2}$$

Por lo tanto, el $R^2 \in [0, 1]$, es decir, se trata de un estadístico de fácil interpretación. Si está más cerca del cero el modelo tiene un bajo grado de ajuste, mientras que si está cerca de uno, el modelo posee un mayor grado de ajuste.

En términos matriciales, se tiene lo siguiente. Note que la variable dependiente puede descomponerse en la suma de dos elementos, la estimación o proyección, y el error de estimación:

$$Y = X\hat{\beta} + \hat{u} = \hat{Y} + \hat{u}$$

Restando el valor promedio de la variable dependiente a ambos lados:

$$Y - \bar{Y} = \hat{Y} - \bar{Y} + \hat{u}$$

$$Y - \bar{Y} = (X - \bar{X})\hat{\beta} + \hat{u}$$

Intuitivamente, la regresión ajustaría bien si las desviaciones de Y respecto de su media se explicaran mayormente por las desviaciones de X respecto de su media más que por los residuos. Se define la siguiente matriz:

$$M^0 = I - \frac{1}{n}ii'$$

que al premultiplicar una matriz la transforma en desvíos con respecto a su media. El vector i es un vector columna de unos. Por lo tanto, se puede transformar el modelo en desvíos con respecto a la media premultiplicándolo por M^0:[29]

$$M^0 y = M^0 X\hat{\beta} + M^0\hat{u} \tag{39}$$

[29] Esta matriz es idempotente y simétrica. A continuación se muestra cómo M^0 transforma un vector en desvíos con respecto a la media. En efecto, note lo siguiente:

$$\begin{pmatrix} \bar{X}_j \\ \bar{X}_j \\ \vdots \\ \bar{X}_j \end{pmatrix} = i\bar{X}_j = i\frac{1}{n}i'X_j = \frac{1}{n}ii'X_j$$

La expresión anterior corresponde a un vector en donde cada uno de sus elementos es el valor promedio del j-ésimo regresor. Luego, es posible escribir el vector j-ésimo en desviaciones respecto

Luego, la variación total de y (la suma de los cuadrados totales) se obtiene premultiplicando la expresión (39) por su traspuesta:

$$STC = (M^0y)'(M^0y) = (M^0X\hat{\beta} + M^0\hat{u})'(M^0X\hat{\beta} + M^0\hat{u})$$

Luego, el lado izquierdo de la expresión queda de la siguiente manera:

$$(M^0y)'(M^0y) = y'M^{0\prime}M^0y = y'M^0y$$

ya que M^0 es idempotente. Finalmente:

$$y'M^0y = \hat{\beta}'X'M^0X\hat{\beta} + \hat{u}'\hat{u}$$

De esta forma, la suma de cuadrados totales (STC) es igual a la suma explicada de los cuadrados (SEC) más la suma de los cuadrados de los errores (SCE). Por lo tanto, una medida respecto de qué tan bien la regresión ajusta a los datos viene dada por:

$$\frac{SSR}{SST} = \frac{\hat{\beta}'X'M^0X\hat{\beta}}{y'M^0y} = 1 - \frac{\hat{u}'\hat{u}}{y'M^0y} \tag{40}$$

Note que al multiplicar y dividir la segunda parte de la expresión anterior por $1/n$ se obtiene:

$$R^2 = 1 - \frac{\hat{u}'\hat{u}(1/n)}{y'M^0y(1/n)} = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_y^2}$$

Al R^2 se le conoce como el coeficiente de determinación. Esta medida se encuentra entre cero y uno, y mide la proporción de la variación total de y que es explicada por la variación de los regresores. Toma el valor cero si la regresión es una línea horizontal, esto es, si todos los elementos de $\hat{\beta}$ excepto la constante son cero. En ese caso, X no explicaría la variabilidad de y. Por otro lado, R^2 es igual a uno si los valores de X e y están todos en la misma línea (hiperplano), por lo que los residuos serían cero.

Varias cosas deben mencionarse respecto de esta medida de ajuste. El R^2 es una medida muy práctica pues al encontrarse entre cero y uno, es de muy fácil

---

a la media, a través de la siguiente expresión:

$$\begin{pmatrix} X_{1j} - \bar{X}_j \\ X_{2j} - \bar{X}_j \\ \vdots \\ X_{nj} - \bar{X}_j \end{pmatrix} = (X_j - i\bar{X}_j) = \left(X_j - \frac{1}{n}ii'X_j\right)$$

Luego:

$$\left(X_j - \frac{1}{n}ii'X_j\right) = \left(I - \frac{1}{n}ii'\right)X_j = M^0X_j$$

Con este procedimiento se ha transformado el vector X_j en desvíos respecto de su media. Si se quisiera transformar toda la matriz X en desvíos respecto de la media, entonces, habría que calcular M^0X.

interpretación.[30] Un elevado R^2 indica un buen grado de ajuste del modelo. Note que dado que el estimador MCO minimiza la suma de los errores al cuadrado, entonces, automáticamente maximiza el valor para R^2. Como dato adicional, cabe señalar que en datos de series de tiempo, el R^2 es elevado debido a la tendencia común que exhiben la mayoría de las series, lo que hace que tiendan a moverse juntas en el tiempo, generando así un elevado valor para este estadístico. En cambio, en datos de corte transversal el R^2 no es muy alto, y de hecho, un valor de 30 % se considera aceptable.

Una nota de precaución, enfocarse sólo en obtener un elevado R^2 no es muy conveniente, por cuanto se estimarán parámetros hechos a la medida para la muestra pero que no capturarán de buena manera la verdadera relación existente entre las variables de interés. Por lo tanto, un elevado R^2 no quiere decir necesariamente que se está en presencia de una buena estimación. De hecho, la principal desventaja del R^2 es que es monotónico, es decir, aumenta a medida que se introducen más regresores al modelo.[31] No obstante, esto no es gratis, pues con ello se perderá precisión en las estimaciones, es decir, los coeficientes estimados tendrán una mayor varianza.

Es por todo lo anterior que los econometristas ponen más atención en el R^2 ajustado ($\bar{R}^2$), el cual se define de la siguiente manera:[32]

$$\bar{R}^2 = 1 - \frac{\hat{u}'\hat{u}/(n-k)}{Y'M^0Y/(n-1)} \tag{41}$$

Es posible mostrar que :

$$\bar{R}^2 = 1 - \frac{SCE/(n-k)}{STC/(n-1)}$$

$$\bar{R}^2 = 1 - (1 - R^2)\left(\frac{n-1}{n-k}\right) \tag{42}$$

La principal ventaja del $\bar{R}^2$ es que no es monotónico, es decir, podría disminuir al incluir un regresor adicional al modelo; incluso podría llegar a ser negativo.

3.4.1. Caso particular: bondad de ajuste de la regresión en el modelo de dos variables

Considere el siguiente modelo que ha sido estimado por MCO:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

[30] Si el modelo incluye una constante (β_1), siempre se cumplirá lo siguiente: $0<R^2<1$. Sin embargo, si el modelo no incluyera una constante, entonces, se obtendrán estimaciones para R^2 fuera de este rango.

[31] Es posible demostrar lo siguiente:

$$R^2_{XZ} = R^2_X + (1 - R^2_X) r^{*2}_{yZ}$$

donde R^2_{XZ} es el R^2 del modelo en donde se ha añadido un nuevo regresor (Z), R^2_X es el R^2 del modelo original, y r^{*2}_{yZ} es el coeficiente de correlación parcial (al cuadrado) entre y y Z, controlando por X. Así, al incluir un regresor adicional, el R^2 del modelo inevitablemente aumentará.

[32] Su nombre se debe a que este estadístico ajusta por la pérdida de grados de libertad en que se incurre al incluir más regresores en la estimación del modelo.

Aplicando sumatoria y dividiendo por n la expresión anterior:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2\bar{X} + \bar{\hat{u}}$$

Restando ambas expresiones:

$$Y_i - \bar{Y} = \hat{\beta}_2(X_i - \bar{X}) + \hat{u}_i$$

Luego:

$$y_i = \hat{\beta}_2 x_i + \hat{u}_i$$

Por lo tanto:

$$\hat{u}_i = y_i - \hat{\beta}_2 x_i$$

Elevando al cuadrado esta expresión:

$$\hat{u}_i^2 = y_i^2 - 2y_i\hat{\beta}_2 x_i + \hat{\beta}_2^2 x_i^2$$

Aplicando sumatoria:

$$\sum \hat{u}_i^2 = \sum y_i^2 - 2\hat{\beta}_2 \sum y_i x_i + \hat{\beta}_2^2 \sum x_i^2$$

Pero se sabe que:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

Por lo que:

$$\sum x_i y_i = \hat{\beta}_2 \sum x_i^2$$

Reemplazando:

$$\sum \hat{u}_i^2 = \sum y_i^2 - 2\hat{\beta}_2(\hat{\beta}_2 \sum x_i^2) + \hat{\beta}_2^2 \sum x_i^2$$

Simplificando:

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2$$

Luego:

$$\sum y_i^2 = \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2$$

Finalmente:

$$R^2 = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2}$$

3.5. Inferencia en el contexto de MCO

La inferencia estadística en el contexto del análisis de regresión es muy útil para testar distintos modelos teóricos. A modo de ejemplo, considere el siguiente modelo que indica que la inversión (I) es función de la tasa de interés nominal (i) y de la inflación (π):

$$I_t = \beta_1 + \beta_2 i_t + \beta_3 \pi_t + u_t$$

Es posible establecer que la inversión, al constituir una decisión que tiene resultados a largo plazo, depende más bien de la tasa de interés real (r). Luego, se podría reformular la especificación anterior de la siguiente manera:

$$I_t = \beta_1 + \beta_2 (i_t - \pi_t) + u_t$$

en donde se ha impuesto la siguiente restricción sobre los parámetros: $\beta_2 + \beta_3 = 0$.[33] Así, haciendo un test de hipótesis sobre los coeficientes estimados ($\hat{\beta}_2$ y $\hat{\beta}_3$) se podría corroborar esta hipótesis. En términos más generales, es posible señalar que la inferencia estadística tiene los siguientes objetivos:

- Testar la significancia individual de los coeficientes estimados
- Construir intervalos de confianza
- Testar hipótesis de interés teórico; por ejemplo, en una función de producción Cobb-Douglas se podría testar la existencia de rendimientos constantes a escala
- Testar la significancia global del modelo, es decir, que todos los coeficientes del modelo, salvo la constante, son cero

Antes de considerar cada uno de los objetivos anteriores, es necesario aclarar dos cosas. En primer lugar, en lo que sigue sólo se realizarán test de hipótesis lineales sobre los coeficientes. Y en segundo lugar, para todos los estadísticos que se analizarán es esencial el supuesto de normalidad del término de error (shock estocástico). Es importante recordar algunas definiciones importantes sobre la inferencia estadística. El error de tipo I, también conocido como nivel de significancia del test o tamaño del test, es la probabilidad de rechazar la hipótesis nula cuando la hipótesis nula es verdadera. Al error de tipo I se le denomina α. El error de tipo II es la probabilidad de no rechazar la hipótesis nula cuando la hipótesis nula es falsa. A este error se le denomina β. El poder del test, que es igual a 1-β, es la probabilidad de rechazar la hipótesis nula cuando la hipótesis nula es falsa. Por lo tanto, se hace evidente que es conveniente trabajar con test de bajo tamaño (α) y elevado poder (bajo β). Lo que se hace en la práctica es escoger α (1, 5 o 10 por ciento) y trabajar con test de buen poder.

[33] Recuerde que:

$$r \approx i - \pi$$

3.5.1. Test de hipótesis sobre los coeficientes individuales

Sea S^{jj} el j-ésimo elemento de la diagonal de la matriz $(X'X)^{-1}$. Luego, asumiendo normalidad en el término de error, el siguiente término:

$$z_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\sigma^2 S^{jj}}} \sim N(0,1) \tag{43}$$

tiene una distribución normal estándar. Si σ^2 fuera conocido, la inferencia estadística en torno a β_j se basaría en el estadístico z_j. Sin embargo, al utilizar $\tilde{\sigma}^2$ en lugar de σ^2 se debe derivar un estadístico alternativo.

Considere la siguiente expresión:

$$\frac{(n-k)\tilde{\sigma}^2}{\sigma^2} = \frac{\hat{u}'\hat{u}}{\sigma^2} = \left(\frac{u}{\sigma}\right)' M \left(\frac{u}{\sigma}\right) \tag{44}$$

la cual se distribuye chi-cuadrado con $n-k$ grados de libertad. En efecto, la suma de n variables aleatorias estandarizadas y elevadas al cuadrado se distribuye chi-cuadrado con n grados de libertad. En este caso, como ya han sido estimados k parámetros, entonces, tiene $n-k$ grados de libertad. Más formalmente, es posible demostrar lo anterior a partir de la siguiente propiedad: si A es una matriz simétrica e idempotente de rango d, entonces:

$$x'Ax \sim \chi^2(d)$$

Por lo tanto:

$$u'Mu \sim \chi^2(r)$$

donde r es el rango (ρ) de la matriz M, la cual es idempotente y simétrica. Dado lo anterior, es posible aplicar otra propiedad que establece que el rango de una matriz idempotente y simétrica es igual a la traza (tr). Luego:

$$\rho(M) = tr[I_n - X(X'X)^{-1}X'] = n - tr[(X'X)^{-1}X'X] = n-k$$

Finalmente, dado que $\hat{u}'\hat{u} = u'Mu$, entonces:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2_{n-k}$$

Luego, el cuociente:

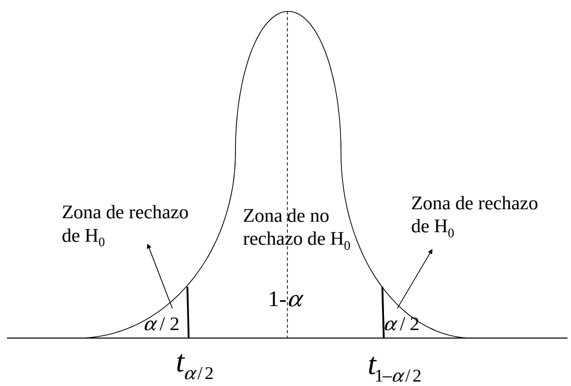
$$t_j = \frac{(\hat{\beta}_j - \beta_j)/\sqrt{\sigma^2 S^{jj}}}{\sqrt{[(n-k)\tilde{\sigma}^2/\sigma^2]/(n-k)}} = \frac{\hat{\beta}_j - \beta_j}{\sqrt{\tilde{\sigma}^2 S^{jj}}} \tag{45}$$

se distribuye t de student con $(n-k)$ grados de libertad.[34] Es posible utilizar el estadístico t_j para testar hipótesis (cuando hay una sola hipótesis a testear) o construir intervalos de confianza de elementos individuales del vector de parámetros β.

[34] La distribución t de student se obtiene a partir del cuociente entre una variable aleatoria normal estandarizada y la raíz cuadrada de una variable chi-cuadrado dividida por sus grados de libertad.

Un test muy utilizado es el denominado test t de significancia estadística, que consiste en determinar si el coeficiente estimado es estadísticamente distinto de cero ($H_0 : \beta_j = 0$). En términos conceptuales este test evalua la probabilidad que el coeficiente estimado sea estadísticamente igual a cero, tal como lo muestra la figura 9.

Figura 9: Significancia estadística



Luego, el test t para evaluar la significancia individual de $\hat{\beta}_j$ viene dado por:

$$t^c = \frac{\hat{\beta}_j}{\sqrt{var(\hat{\beta}_j)}} \tag{46}$$

donde la hipótesis nula es que el coeficiente es cero. Si $t^c > t_{1-\alpha/2}$, donde $t_{1-\alpha/2}$ es el valor crítico de la distribución t con $n-k$ grados de libertad, entonces, se rechaza la hipótesis nula y el coeficiente es estadísticamente significativo, es decir, distinto de cero. El valor crítico $1,96$ es el que se usa habitualmente para determinar la significancia estadística de un coeficiente en un análisis de regresión.[35]

Es necesario hacer algunos comentarios con respecto al test t. En primer lugar, se debe señalar que enfocarse excesivamente en el test t para decidir si incluir o no una determinada variable en el modelo puede inducir a errores, ya que el hecho de que un coeficiente no sea estadísticamente significativo no quiere decir que no sea económicamente significativo. Es por esto que se debe tener precaución con el test t, y no asignarle más valor del que realmente tiene. Por otro lado, se debe notar que el valor del t calculado es muy sensible al tamaño de la muestra (n). De hecho, si esta es muy grande, entonces, la varianza del estimador cae, y por ende, el valor del test t aumenta. De esta manera, se obtendrán test t muy grandes, por lo que los coeficientes estimados serán con una alta probabilidad significativos, ya que será

[35] Por lo general, los test se hacen al 5 %, y la razón de ello no es clara. Solo es posible mencionar que en un artículo muy citado Fisher (1923) sugiere utilizar este valor, y hoy todos lo usan, respondiendo a una especie de acuerdo social, pues no hay ninguna razón de peso para inclinarse a favor del 5 % en lugar del 6 o de un 7 %.

muy probable que el valor del test se encuentre en la zona de rechazo de la hipótesis nula.[36]

3.5.2. Intervalo de confianza para los coeficientes estimados

Un intervalo de confianza para el coeficiente $\hat{\beta}_j$ se puede construir a partir del test t. En efecto:

$$t = \frac{\hat{\beta}_j - \beta_j}{\sqrt{var(\hat{\beta}_j)}} \sim t_{n-k}$$

El nivel de confianza de este intervalo (1-α) por lo general es de 90, 95 o 99 por ciento. Luego, lo que se hace es construir la probabilidad de que el valor del test t se encuentre en una determinada región de tamaño 1-α:

$$P\left(t_{\alpha/2,n-k} \leq \frac{\hat{\beta}_j - \beta_j}{\sqrt{var(\hat{\beta}_j)}} \leq t_{1-\alpha/2,n-k}\right) = 1 - \alpha$$

dado que la distribución t es simétrica, entonces:

$$P\left(-t_{1-\alpha/2,n-k} \leq \frac{\hat{\beta}_j - \beta_j}{\sqrt{var(\hat{\beta}_j)}} \leq t_{1-\alpha/2,n-k}\right) = 1 - \alpha$$

Desarrollando la expresión anterior se llega a lo siguiente:

$$P(\hat{\beta}_j - t_{1-\alpha/2,n-k}\sqrt{var(\hat{\beta}_j)} < \beta_j < \hat{\beta}_j + t_{1-\alpha/2,n-k}\sqrt{var(\hat{\beta}_j)}) = 1 - \alpha \qquad (47)$$

donde $1 - \alpha$ corresponde al nivel deseado de confianza y $t_{\alpha/2,n-k}$ es el valor crítico de la distribución t con $n - k$ grados de libertad.

Finalmente, es posible construir intervalos de confianza para la varianza del término de error (σ^2). Para ello, se parte del siguiente hecho:

$$\frac{\hat{u}'\hat{u}}{\sigma^2} \sim \chi^2_{n-k}$$

Luego:

$$\frac{(n-k)\tilde{\sigma}^2}{\sigma^2} \sim \chi^2_{n-k}$$

[36] Para realizar un test de una sola cola (lo que corresponde a una hipótesis nula de desigualdad) se ajusta la región crítica y se utiliza el punto crítico de t_α de la distribución. Por ejemplo, para testar la hipótesis nula $H_0 : \hat{\beta}_j = 1$ versus $H_1 : \hat{\beta}_j < 1$ se construye el siguiente estadístico:

$$t = \frac{\hat{\beta}_j - 1}{\sqrt{var(\hat{\beta}_j)}}$$

y se compara con el valor crítico de tabla correspondiente (α). Si el valor del test t calculado es menor que el valor crítico (t_α), entonces, se rechaza la hipótesis nula.

El intervalo de confianza exige que se cumpla lo siguiente:

$$P\left(\chi_{\alpha/2,n-k} \leq \frac{(n-k)\tilde{\sigma}^2}{\sigma^2} \leq \chi_{1-\alpha/2,n-k}\right) = 1-\alpha$$

Luego de un poco de álgebra se llega a lo siguiente:

$$P\left(\frac{(n-k)\tilde{\sigma}^2}{\chi_{1-\alpha/2,n-k}} \leq \sigma^2 \leq \frac{(n-k)\tilde{\sigma}^2}{\chi_{\alpha/2,n-k}}\right) = 1-\alpha \qquad (48)$$

3.5.3. Testando la significancia de la regresión

Un aspecto fundamental del análisis de regresión es determinar si la estimación del modelo es significativa como un todo, es decir, si el modelo es capaz de explicar en conjunto la variación de la variable dependiente. Esto se operacionaliza a través de un test en donde se evalúa si todos los coeficientes estimados, excepto la constante, son estadísticamente significativos. Si todos los coeficientes asociados a los regresores incluidos son estadísticamente iguales a cero, entonces, el coeficiente de correlación múltiple también tenderá a cero, por lo tanto, el test se basa precisamente en el valor del R^2. El siguiente estadístico permite realizar la prueba ya mencionada:

$$F_{k-1,n-k} = \frac{SEC/(k-1)}{SCE/(n-k)}$$

donde SEC es la suma explicada de los cuadrados, y SCE es la suma de los cuadrados de los errores. Si no se rechaza la hipótesis nula de que todos los coeficientes son estadísticamente no significativos, salvo la constante, y además, los errores se distribuyen normal, entonces, el estadístico tiene una distribución F con $k-1$ y $n-k$ grados de libertad en el numerador y denominador, respectivamente. Es posible mostrar que la expresión anterior es equivalente a:

$$F_{k-1,n-k} = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \qquad (49)$$

Así, valores elevados para F (un alto R^2) proporcionan evidencia en contra de la validez de la hipótesis nula. En concreto, si el valor F calculado es mayor que el valor crítico, entonces, se rechaza la hipótesis nula.

Por otro lado, si se quisiera testar que solo un subconjunto de las pendientes del modelo (por ejemplo, $\hat{\beta}_4$, $\hat{\beta}_6$, y $\hat{\beta}_{12}$) son estadísticamente iguales a cero, entonces, se puede utilizar la siguiente especificación para el test F:

$$F_{k_2,n-k} = \frac{(\hat{u}_*'\hat{u}_* - \hat{u}'\hat{u})/k_2}{\hat{u}'\hat{u}/(n-k)}$$

donde u_* corresponde a los residuos de la estimación restringida, es decir, se obtienen estimando el modelo original pero imponiendo como cierta la hipótesis nula (en este caso, asumiendo que $\hat{\beta}_4 = \hat{\beta}_6 = \hat{\beta}_{12} = 0$). k_2 corresponde la diferencia en los grados de libertad entre el modelo restringido y el modelo no restringido. Si el valor

F calculado es mayor que el valor crítico, entonces, se rechaza la hipótesis nula. Alternativamente, el test anterior puede escribirse como:

$$F_{k_2,n-k} = \frac{(R^2 - R_*^2)/k_2}{(1-R^2)/(n-k)}$$

3.5.4. Significancia individual versus significancia global del modelo

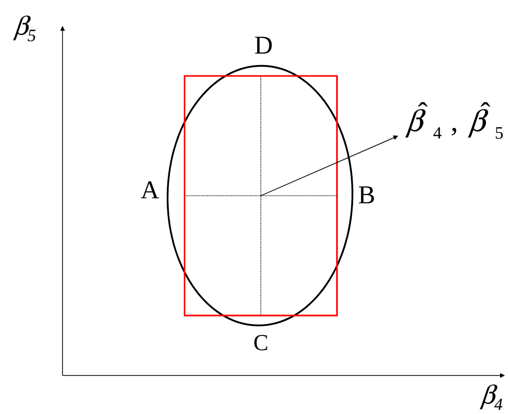
Considere el siguiente modelo:

$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i} + u_i$$

con $i = 1, 2, ..., n$. Considere una situación en donde tanto $\hat{\beta}_4$ como $\hat{\beta}_5$ no son estadísticamente significativos a nivel individual. Un aspecto interesante de notar es que, a pesar de lo anterior, es posible que en conjunto ambos coeficiente sí sean estadísticamente significativos, lo cual se evalúa por medio de un test F. Aquello podría darse por la covarianza existente entre los coeficientes estimados. En efecto, es posible que el área roja de la figura 8 sea significativa en cuanto a tamaño, y que la covarianza existente entre ambos coeficientes sea tal que en conjunto logren explicar la variabilidad de la variable dependiente.

Considere la estimación de MCO para $\hat{\beta}_4$ y $\hat{\beta}_5$, que viene dada por el punto del centro del rectángulo de la figura 10. Los trazos AB y CD representan los intervalos de confianza para cada una de las estimaciones, por ejemplo, al 95 %.[37] Un intervalo de confianza para el vector $\hat{\beta}_4$ y $\hat{\beta}_5$ viene dado por el área que, en repetidas muestras, cubre el valor en, por ejemplo, el 95 % de las ocasiones.

Figura 10: Significancia individual versus significancia global

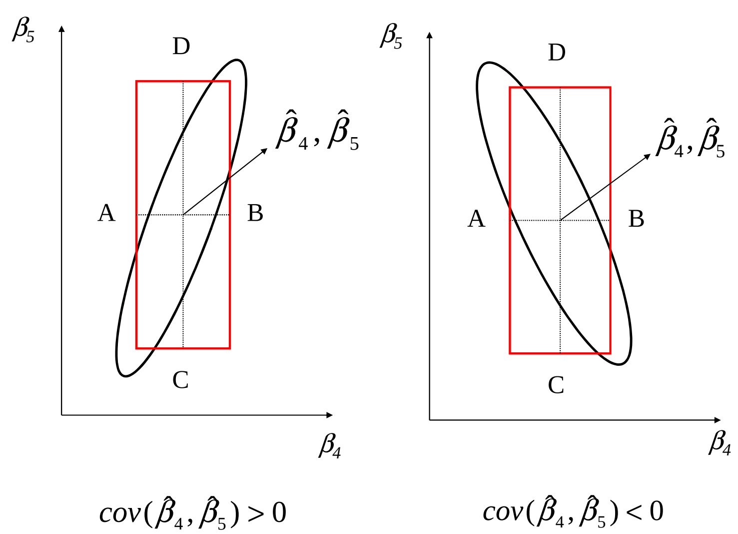


Por lo tanto, un candidato natural para construir dicha área consiste en el rectángulo formado por las estimaciones de los intervalos individuales. Si la covarianza de las estimaciones de $\hat{\beta}_4$ y $\hat{\beta}_5$ es nula (es decir, si ambos estimadores son independientes), entonces, en repetidas muestras este rectángulo contendrá al punto desconocido

[37] Note que se ha asumido que $\hat{\beta}_5$ tiene un error estándar mayor que $\hat{\beta}_4$.

($\hat{\beta}_4$ y $\hat{\beta}_5$) el 95 %· 95 % de las veces, esto es, en el 90,25 % de los casos. Por lo tanto, este rectángulo no es tan grande para construir un intervalo de confianza del 95 %, se requiere entenderlo. ¿De dónde se podría agrandar este rectángulo? Debería ampliarse en aquellas áreas en donde existe una alta probabilidad de cubrir ($\hat{\beta}_4$ y $\hat{\beta}_5$), es decir, que se den aquellos valores. Note que las esquinas del rectángulo tienen una baja probabilidad de ocurrencia ya que los valores estimados están muy lejos de sus valores medios. Por lo tanto, las áreas que están justo fuera del rectángulo, pero cerca de los puntos A, B, C y D, tienen una mayor probabilidad asociada de cubrir el punto ($\hat{\beta}_4$,$\hat{\beta}_5$) que las áreas de las esquinas. El test F permitirá construir precisamente el intervalo de confianza como una elipse (ya que hay dos coeficientes involucrados), que es la que se presenta en la figura. La forma que tiene la elipse de la figura 10 corresponde a un caso en donde la covarianza entre los coeficientes estimados es nula, la figura 11 ilustra el caso en donde se tiene una covarianza positiva y una negativa entre los coeficientes $\hat{\beta}_4$ y $\hat{\beta}_5$. Esta última (covarianza negativa) estaría reflejando el hecho que sobrestimaciones de un parámetro provocan subestimaciones del otro. Note además que siempre la elipse se encuentra centrada en el punto ($\hat{\beta}_4$,$\hat{\beta}_5$).

Figura 11: Significancia individual versus significancia global: las covarianzas



Este ejemplo ilustra el caso en que los coeficientes podrían no ser estadísticamente significativos a nivel individual, pero en conjunto sí serlos. En efecto, suponga que se testean las hipótesis nulas de que los coeficientes estimados ($\hat{\beta}_4$ y $\hat{\beta}_5$) son individualmente no significativos. Si el punto (0,0) permanece en una esquina del rectángulo, los coeficientes efectivamente no serán estadísticamente significativos. Por lo tanto, los test t no rechazan la hipótesis nula. Sin embargo, al testear la hipótesis de significancia conjunta de los coeficientes, el test F concluye que los coeficientes estimados son estadísticamente significativos, pues el punto (0,0) permanece fuera de la elipse (recuerde que está en una esquina del rectángulo). Es decir, al menos una de las variables tiene una influencia significativa sobre la variable dependiente, pero no se puede establecer cuál. Note, por la forma de las elipses, que esta situación es más probable que se dé en situaciones donde existe una alta correlación (negativa o positiva) entre las variables independientes.

3.6. Testando un conjunto de restricciones lineales

Considere el siguiente modelo de regresión lineal:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

En este contexto, podría ser interesante testar no una sino un conjunto de restricciones (hipótesis) lineales. Sin importar cuál sea el número de las restricciones a testar o los coeficientes involucrados, el conjunto de restricciones siempre puede escribirse, en términos matriciales, de la siguiente forma:

$$R\beta = r \tag{50}$$

donde R es una matriz de qxk, β es el vector columna de coeficientes de dimensión k, y r es un vector columna de constantes de dimensión q. Así, cada una de las filas de R contiene los escalares asociados a cada una de las restricciones a testar. La matriz R tiene k columnas (número de parámetros), q filas (número de hipótesis), y rango columna completo, por lo que q debe ser menor que k. Además, las filas de R deben ser linealmente independientes. Considere los siguientes ejemplos de hipótesis lineales a testar y note la forma que tendría la matriz R en cada uno de estos casos:

- Uno de los coeficientes es cero ($\beta_j = 0$):

$$R = [0 \;\; 0 \cdots 1 \;\; 0 \cdots \;\; 0] \;\; con \; r = 0$$

- Dos de los coeficientes son iguales ($\beta_2 = \beta_3$):

$$R = [0 \;\; 1 \;\; -1 \;\; 0 \cdots 0] \;\; con \; r = 0$$

- Un conjunto de coeficientes suman uno ($\beta_2 + \beta_3 + \beta_4 = 1$):

$$R = [0 \;\; 1 \;\; 1 \;\; 1 \;\; 0 \cdots 0] \;\; con \; r = 1$$

- Un subconjunto de los coeficientes son iguales a cero ($\beta_1 = \beta_2 = \beta_3 = 0$):

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \end{pmatrix}$$

  con:

$$r = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

  o equivalentemente:

$$(I_3 : 0)\beta = 0$$

- Varias restricciones simultáneas ($\beta_2 + \beta_3 = 1$, $\beta_4 + \beta_6 = 0$, $\beta_5 + \beta_6 = 2$):

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \\ \beta_6 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

Por lo tanto, note que sin importar el tipo de hipótesis a testar, estas siempre se podrán escribir como $R\beta = r$, con R siendo una matriz de constantes, cuyo número de filas corresponde al número de hipótesis lineales a testar, y r un vector de constantes.

A continuación se asumirá que el término de error (u_i) tiene una distribución normal. Considere un conjunto de q hipótesis lineales de la forma:

$$H_0 : R\beta = r$$

Cabe señalar, que en la mayoría de los casos la matrix R tendrá pocos unos (1) y varios ceros (0), dependiendo de cuáles sean los coeficientes que entran en la hipótesis a testear. La hipótesis, entonces, debería centrarse en las discrepancias que exhiba el vector $m = R\hat{\beta} - r$. La cuestión estadística se refiere al hecho de si la desviación de m respecto al cero puede ser atribuida a un error muestral, o bien, a una desviación significativa. Dado que $\hat{\beta}$ se distribuye normal (porque u se distribuye normal) y m es una función lineal de $\hat{\beta}$, entonces, m también se distribuye normal. Si la hipótesis nula es cierta, m tiene media vectorial cero[38] y su varianza viene dada por:

$$var(m) = var(R\hat{\beta} - r) = var(R\hat{\beta}) = R[var(\hat{\beta})]R' = \sigma^2 R(X'X)^{-1}R'$$

De esta manera, el test podría basarse en el criterio de Wald, el cual establece lo siguiente:

$$W = m'[var(m)]^{-1}m \sim \chi^2(q)$$

El estadístico W tiene una distribución chi-cuadrado con q grados de libertad si la hipótesis nula es correcta. Sin embargo, el estadístico chi-cuadrado no es aplicable en este caso pues se desconoce el término σ^2, y hay que estimarlo. Sin embargo, es posible utilizar un estadístico similar al test t por medio de incluir $\tilde{\sigma}^2$ en lugar de σ^2:[39]

$$F = \frac{(R\hat{\beta} - r)'[\sigma^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/q}{[(n-k)\tilde{\sigma}^2/\sigma^2]/(n-k)}$$

Cancelando términos la expresión queda de la siguiente manera:

$$F = \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)/q}{\tilde{\sigma}^2}$$

[38]En efecto, si la hipótesis nula es cierta, entonces, $R\beta = r$, y dado que $E(R\hat{\beta}) = R\beta$, entonces, $E(m) = 0$.

[39]Una distribución F se genera a partir del cuociente entre dos variables chi-cuadrado divididas cada una por sus respectivos grados de libertad.

Finalmente:

$$F = \frac{(R\hat{\beta} - r)'[\tilde{\sigma}^2 R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)}{q} \sim Fq, n-k \tag{51}$$

Esta es la formulación más general para el test F. Recuerde que cuando se implementa el test para evaluar la significancia conjunta del modelo esta expresión se reduce al test F que se vió anteriormente.

En el caso particular que se trate de una sola hipótesis a testar, por ejemplo $\beta_j = r$, la expresión anterior se reduce a:

$$\frac{(\hat{\beta}_j - r)^2}{var(\hat{\beta}_j)} \sim F(1, n-k)$$

Se sabe además que una variable aleatoria F con un grado de libertad en el numerador y n grados de libertad en el denominador se distribuye t_n^2. Por lo tanto, la raíz cuadrada a la expresión anterior se distribuye t, es decir:

$$\frac{\hat{\beta}_j - r}{\sqrt{var(\hat{\beta}_j)}} \sim t_{n-k}$$

En el caso que se trate de un test de significancia individual, se tiene que $r = 0$, con lo cual se llega:

$$\frac{\hat{\beta}_j}{\sqrt{var(\hat{\beta}_j)}} \sim t_{n-k}$$

que corresponde al test t ya presentado. En el caso que la hipótesis nula a testar sea de la forma $H_0 : \beta_i + \beta_j = 1$, entonces, el test F se reduce a:

$$\frac{\hat{\beta}_i + \hat{\beta}_j - 1}{\sqrt{var(\hat{\beta}_i) + var(\hat{\beta}_j) + 2cov(\hat{\beta}_i, \hat{\beta}_i)}} \sim t_{n-k}$$

3.6.1. El p-value

El p-value es el nivel de significatividad más pequeño al que se rechazaría la hipótesis nula. El p-value es la probabilidad de observar un estadístico (el test t por ejemplo) tan extremo como el que se ha obtenido si es que la hipótesis nula es verdadera. Por lo tanto, el p-value refleja la fuerza o debilidad de la evidencia empírica frente a la hipótesis nula. Luego, un p-value pequeño representa evidencia en contra de la hipótesis nula, y un p-value grande constituye evidencia a favor.

A modo de ejemplo, considere el test t de dos colas para construir el p-value (el test de significancia individual):

$$p = P(|t^*| \geq |t^c|) = 2(1 - \Phi(|t^c|))$$

donde Φ representa la función de distribución acumulada de una normal estándar. El p-value es reportado de manera automática por los distintos paquetes econométricos al momento de hacer una estimación. Por ejemplo, EVIEWS al lado de cada coeficiente estimado presenta el p-value del test de significancia individual del coeficiente. Por ejemplo, si el p-value de un determinado coeficiente es igual a 0,04 quiere decir que la hipótesis nula (en este caso que el coeficiente estimado sea igual a cero) no es significativa al 5 por ciento. Sin embargo, la hipótesis nula sí es significativa al 1 por ciento. En efecto, recuerde que el p-value es el nivel de significatividad más pequeño al que se rechazaría la hipótesis nula, en este caso ese nivel de significatividad es del 4 por ciento. Por ende, si la hipótesis nula se rechaza al 4 por ciento, con más razón se rechaza al 5 por ciento, que es un nivel menos restrictivo. Sin embargo, no es posible rechazarla al 1 por ciento en vista de la evidencia empírica proporcionada. Dicho de otro modo, suponga que usted desea tener coeficientes que sean estadísticamente signficativos al 5 por ciento, es decir, $\alpha = 5$ por ciento. Como ese es el nivel de signficatividad más pequeño que usted está requiriendo para sus coeficientes, entonces, los coeficientes estadísticamente significativos serán aquellos cuyo p-value sea menor que el α escogido. Note que el contar con el p-value ahorra el hecho de tener que contar con tablas estadísticas para poder evaluar la signficancia de los parámetros estimados, o de cualquier otro test de hipótesis que se quiera llevar a cabo. Por ende, es del todo práctico contar con el p-value para los test de hipótesis que se quiera realizar.

La tabla 2 presenta los resultados de una ecuación de Mincer que incluye además de la escolaridad, la edad y la edad al cuadrado como variables explicativas.[40]

Tabla 2: El modelo de regresión lineal general
Dependent Variable:LWPH
Method: Least Squares
Sample: 12900
Included observations: 2900

Variable	**Coefficient**	**Std.Error**	**t-Statistic**	**Prob.**
C	5.430366	0.111611	48.65426	0.0000
ESC	0.092234	0.002843	32.43769	0.0000
EDAD	0.023937	0.005645	4.240247	0.0000
EDAD2	-0.000133	7.04E-05	-1.890245	0.0588
R-squared	0.275247	Mean dependent var		7.112859
Adjusted R-squared	0.274496	S.D dependent var		0.649708
S.E of regression	0.553399	Akaike info criterion		1.655902
Sum squared resid	886.8997	Schwarz criterion		1.664140
Log likelihood	-2397.057	Hannan-Quinn criter.		1.658870
F-statistic	366.6141	Durbin-Watsons stat		2.059148
Prob (F-statistic)	0.000000			

La estimación por mínimos cuadrados ordinarios revela que por cada año adicional de escolaridad el salario por hora crece en promedio en un 9.22 %. Por otro lado,

[40]La edad y la edad al cuadrado se agregan al modelo como de manera de incorporar una medida de la experiencia laboral de la persona. El término cuadrático intenta capturar un perfil no lineal en los retornos a la edad del individuo.

el retorno a la edad cóncavo (crece pero a tasas decrecientes) ya que el coeficiente que acompaña a la variable cuadrática es negativo. Los test t muestran que todos los coeficientes, salvo el de la edad al cuadrado, son estadísticamente significativos (distintos de cero). El p-value confirma lo anterior, pero además muestra que los coeficientes son estadísticamente significativos al 1 %. El R^2 y el $\bar{R}^2$ se ubican en torno al 27 %, cifra no menor considerando que se está trabajando con datos de corte transversal.[41]. El p-value del test F revela que el modelo es globalmente significativo, es decir, todas las pendientes, en conjunto, son estadísticamente distintas de cero.

3.6.2. Normalidad de los errores

Tal como ha quedado establecido, para realizar inferencia estadística es muy importante que los errores tengan una distribución normal. Es por ello que se hace necesario testar la distribución de los residuos estimados ($\hat{u}$) para verificar si se cumple o no este supuesto. No obstante lo anterior, aún cuando no se cumpliera es posible justificar la inferencia sobre la base del teorema central del límite.

En primer lugar, se debe tener presente la forma que tiene la distribución normal, es simétrica y mesocúrtica (colas livianas), lo que guarda relación con el tercer y cuarto momento de la distribución. La distribución es simétrica cuando el tercer momento estandarizado es igual a cero, esto es:[42]

$$S = \frac{(E(u - E(u))^3}{(var(u))^{3/2}} = \frac{E(u^3)}{(var(u))^{3/2}} = 0$$

Por otro lado, que una distribución sea mesocúrtica implica que la kurtosis (K), esto es, el cuarto momento de la distribución estandarizado, es igual a tres:

$$K = \frac{E(u - E(u))^4)}{(var(u))^2} = \frac{E(u^4)}{(var(u))^2} = 3$$

El test de Jarque-Bera (1980) testa de manera conjunta el cumplimiento de estos dos aspectos de la siguiente manera:

$$JB = n \left[\frac{\hat{S}}{6} + \frac{(\hat{K} - 3)^2}{24} \right] \sim \chi^2_2$$

La hipótesis nula es de normalidad, por lo tanto, cuando el valor JB supera el valor crítico de tabla de χ^2_2 (5,99) se rechaza la hipótesis nula.

[41]En general en este tipo de datos los R cuadrados suelen ser bajos.

[42]En términos prácticos, suponiendo que se tiene una muestra aleatoria de tamaño n la simetría podría calcularse como:

$$S = \frac{(1/n) \sum (w_i - \bar{w})^{3/2}}{(1/(n-1) \sum (w_i - \bar{w})^2)^(3/2)}$$

3.7. Predicción

No sólo es importante obtener buenas estimaciones de los parámetros poblacionales, sino también poder realizar buenas predicciones con el modelo. Considere el siguiente modelo:

$$y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + u_t$$

Luego, con la estimación MCO puede obtenerse una predicción para y_{t+j} de la siguiente manera:

$$\hat{y}_{t+j} = \beta_1 + \hat{\beta}_2 X_{2,t+j} + \cdots + \hat{\beta}_k X_{k,t+j}$$

Ahora bien, en la práctica se pueden llevar a cabo dos tipos de predicciones:

- Predicción econométrica (modelo econométrico): el modelo se puede utilizar para predecir el valor de la variable dependiente (y) en función de la información existente para las variables independientes (X). Por ende, este tipo de predicción requiere un modelo económico.

- Modelos de series de tiempo: las series de tiempo pueden ser caracterizadas a través de una tendencia (t), un factor estacional, un elemento cíclico y un término de error. En lugar de incluir al tiempo (t) como una variable independiente, se podrían incluir rezagos de la misma variable dependiente (Y_{t-p} con $p \geq 0$). Así, se generan predicciones para el comportamiento de la serie, pero reconociendo que la historia, o el comportamiento pasado influye en la evolución futura de la serie. Este tipo de técnicas se conocen como análisis de Box-Jenkins (modelos ARIMA).[43]

Existen cuatro potenciales fuentes de error al realizar una predicción:

1. Error de especificación: los supuestos no se cumplen (forma funcional, variables incluidas, cambios de régimen).

2. Error de condicionamiento. Suponga que el modelo es el siguiente $y_t = \beta_1 + \beta_2 X_t + u_t$, y se desea hacer una predicción para $t+j$ ($\hat{y}_{t+j}$). Para ello se requiere contar con el valor X_{t+j}, el cual también debe estimarse, y por ende, introduce una fuente de imprecisión en la estimación de $\hat{y}_{t+j}$.

3. Error muestral: se utilizan estimaciones de los parámetros poblacionales para hacer la predicción y no los verdaderos valores.

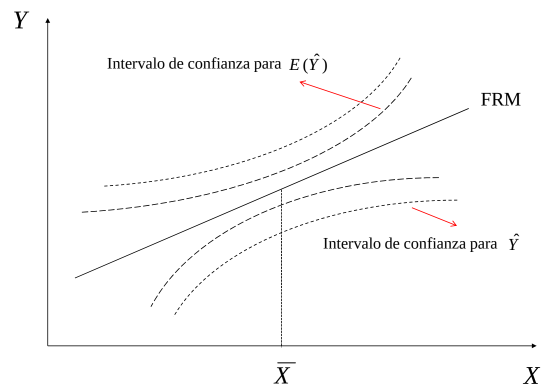
4. Error aleatorio: se asume implícitamente que u_{t+j} es cero, lo cual no es necesariamente cierto. En efecto, la predicción vendría dada por: $\hat{y}_{t+j} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+j} + \hat{u}_{t+j}$. Dado que no se conoce $\hat{u}_{t+j}$ se asume que es cero.

Un aspecto importante de tener presente es el siguiente: para construir el intervalo de confianza de la predicción sólo se consideran los potenciales errores (3) y (4). Si las fuentes de error (1) y (2) no son significativas, entonces, no existe otra

[43] Es directo notar que para datos de corte transversal solo el enfoque econométrico de predicción es válido.

predicción que tenga un intervalo de confianza más eficiente que el de la figura 12.

Figura 12: Intervalo de confianza



Sin embargo, en la práctica eso no ocurre, por lo que los econometristas deben ajustar sus predicciones para dar cuenta de estos fenómenos. En este sentido, las modificaciones de criterio en los modelos econométricos consisten en complementar las predicciones con información de tipo cualitativo y con la experiencia del econometrista, para incorporar efectos de situaciones imprevistas y que afectarán la predicción, como por ejemplo, la ocurrencia de huelgas, cambios de política económica, actos de terrorismo, eventos climáticos, etc.

Suponga que se ha ajustado una regresión a un conjunto de datos, con lo cual se puede obtener la predicción para la observación i-ésima:

$$\hat{y}_i = x_i'\hat{\beta}$$

donde x_i' es un vector fila de dimensión k que tiene la forma: $x_i' = [1 \ X_{2i} \ \cdots \ X_{ki}]$. Suponga ahora que se cuenta con un vector de datos nuevos para x (x_f), y se quiere realizar una predicción, pero además construir un intervalo de confianza para la predicción puntual. Así, la predicción será:[44]

$$\hat{y}_f = x_f'\hat{\beta}$$

Luego, el error de predicción vendrá dado por:

$$\hat{e}_f = y_f - \hat{y}_f\hat{e}_f = X_f\hat{\beta} + u_f - x_f'\hat{\beta}\hat{e}_f = u_f + x_f'(\beta - \hat{\beta})$$

de lo cual se desprende que $E(\hat{e}_f) = 0$. Además, la varianza del error de predicción (necesaria para construir el intervalo de confianza) vendrá dada por:

$$var(\hat{e}_f) = E([x_f'(\beta - \hat{\beta}) + u_f][x_f'(\beta - \hat{\beta}) + u_f]')$$

[44]Toda predicción se realiza bajo el supuesto que el modelo ajustado es válido para el período de predicción, lo cual no necesariamente es cierto. Pudieran ocurrir transformaciones importantes que hagan que el pasado no sea un buen predictor de lo que pasará en el futuro.

$$var(\hat{e}_f) = E[x'_f(\beta - \hat{\beta})(\beta - \hat{\beta})'x_f + 2x'_f(\beta - \hat{\beta})u_f + u_f u'_f]$$

Tomado en cuenta lo siguiente:

$$E[(\beta - \hat{\beta})(\beta - \hat{\beta})'] = \sigma^2 (X'X)^{-1}$$

Por lo tanto, la varianza del error de predicción queda:

$$var(\hat{e}_f) = \sigma^2 x'_f (X'X)^{-1} x_f + \sigma^2$$
$$var(\hat{e}_f) = \sigma^2 (1 + x'_f (X'X)^{-1} x_f)$$

dado que no se conoce la varianza del término de error se utiliza una estimación de éste, por lo tanto, la expresión anterior queda de la siguiente manera:

$$var(\hat{e}_f) = \hat{\sigma}_e^2 = \tilde{\sigma}^2 (1 + x'_f (X'X)^{-1} x_f)$$

Luego, si el término de error del modelo (u) se distribuye normal, entonces, el error de predicción se distribuye::

$$\hat{e}_f \sim N(0, \hat{\sigma}_e^2)$$

Dado que el cuociente entre una variable aleatoria normal estandarizada y la raíz cuadrada de una variable chi-cuadrado dividida por sus grados de libertad se distribuye t, entonces:

$$\frac{\hat{e}_f - E(\hat{e}_f)}{\sqrt{var(\hat{e}_f)}} \sim t_{n-k}$$

Por lo tanto:

$$\frac{y_f - \hat{y}_f}{\sqrt{\tilde{\sigma}^2 (1 + x'_f (X'X)^{-1} x_f)}} \sim t_{n-k}$$

Luego, a partir de la expresión anterior es posible construir un intervalo de confianza para la predicción del valor puntual y_f:

$$P(\hat{y}_f - t_{n-k,1-\alpha/2} \sqrt{\hat{\sigma}_e^2} \leq y_f \leq \hat{y}_f + t_{n-k,1-\alpha/2} \sqrt{\hat{\sigma}_e^2}) = 1 - \alpha$$

Es interesante notar de la figura 12 que el intervalo de confianza alcanza su menor amplitud en el punto medio de las variables explicativas ($\bar{X}$), y a medida que se aleja de dicho punto la predicción tiene asociado un intervalo de confianza más amplio (tanto para la predicción puntual como para la predicción del valor esperado de y). Esto es posible de mostrar a través del modelo más simple. Suponga que el modelo es el siguiente:

$$y_i = \beta_1 + \beta_2 X_i + u_i$$

Anteriormente se derivó la siguiente expresión para la vaianza del error de predicción:

$$\hat{\sigma}_e^2 = \sigma^2 (1 + x'_f (X'X)^{-1} x_f) = \sigma^2 + x'_f \sigma^2 (X'X)^{-1} x_f$$

Dado que se está en presencia del modelo de regresión simple el término x'_f viene dado por $x'_f = (1 \; X_f)$, y la expresión $\sigma^2 (X'X)^{-1}$ corresponde a la matriz de varianzas y covarianzas de β_1 y β_2. Por lo tanto, la ecuación queda de la siguiente forma:

$$\hat{\sigma}_e^2 = \sigma^2 + \begin{pmatrix} 1 & X_f \end{pmatrix} \begin{pmatrix} \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i-\bar{X})^2}\right) & -\frac{\sigma^2\bar{X}^2}{\sum(X_i-\bar{X})^2} \\ -\frac{\sigma^2\bar{X}^2}{\sum(X_i-\bar{X})^2} & \frac{\sigma^2}{\sum(X_i-\bar{X})^2} \end{pmatrix} \begin{pmatrix} 1 \\ X_f \end{pmatrix}$$

Resolviendo la expresión anterior se llega a:

$$\hat{\sigma}_e^2 = \sigma^2\left(1 + \frac{1}{n} + \frac{\bar{X}^2}{\sum(X_i-\bar{X})^2} - \frac{2\bar{X}^2X_f}{\sum(X_i-\bar{X})^2} + \frac{X_f^2}{\sum(X_i-\bar{X})^2}\right)$$

Finalmente:

$$\hat{\sigma}_e^2 = \sigma^2\left(1 + \frac{1}{n} + \frac{(X_f-\bar{X})^2}{\sum(X_i-\bar{X})^2}\right)$$

En esta expresión se aprecia claramente cómo a medida que X_f se aleja de $\bar{X}$ aumenta el error de predicción $\hat{\sigma}_e^2$, y por ende, el intervalo de confianza.

También es posible obtener un intervalo de confianza para el valor esperado de la variable dependiente. Note que en este contexto el error de predicción vendrá dado por:

$$\tilde{e}_f = E(y_f|x_f) - \hat{y}_f$$

Reemplazando términos:

$$\tilde{e}_f = x'_f\beta - x'_f\hat{\beta} = x'_f(\beta - \hat{\beta})$$

Así, la varianza del error de predicción vendrá dada por:

$$var(\tilde{e}_f) = \hat{\sigma}_e^2 = E([x'_f(\beta-\hat{\beta})][x'_f(\beta-\hat{\beta})]') = \tilde{\sigma}^2x'_f(X'X)^{-1}x_f$$

Aplicando el mismo razonamiento utilizado anteriormente, se puede mostrar lo siguiente:

$$\frac{E(y_f|x_f) - \hat{y}_f}{\sqrt{\tilde{\sigma}^2x'_f(X'X)^{-1}x_f}} \sim t_{n-k}$$

Luego, el intervalo de confianza para $E(y)$ vendrá dado por:

$$P(\hat{y}_f - t_{n-k,1-\alpha/2}\sqrt{\hat{\sigma}_e^2} \leq E(y_f|x_f) \leq \hat{y}_f + t_{n-k,1-\alpha/2}\sqrt{\hat{\sigma}_e^2}) = 1-\alpha$$

3.8. Medidas de exactitud de la predicción

Una vez que se ha realizdo una predicción, lo natural es querer evaluarla. Pues bien, a continuación se describen brevemente algunas de las medidas más populares para evaluar la precisión de las predicciones.

Error medio absoluto (MAE). Es apropiado cuando el costo de equivocarse en la predicción es proporcional al tamaño absoluto del error. La expresión de esta medida es la siguiente:

$$MAE = \frac{\sum_{i=1}^{n_f} |y_i - \hat{y}_i|}{n_f}$$

donde n_f corresponde al número de observaciones para las cuales se realizó una predicción, y $|\cdot|$ es la función módulo.

Raíz cuadrada del error cuadrático medio ($RMSE$). Esta medida pondera más fuertemente los grandes errores que los pequeños errores. La expresión de esta medida es la siguiente:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n_f} (y_i - \hat{y}_i)^2}{n_f}}$$

Cabe señalar que la función de pérdida cuadrática es una de las más populares.

Error porcentual absoluto medio ($MAPE$). Corresponde al promedio de los valores absolutos de los errores porcentuales. Es más apropiado cuando el costo de equivocarse depende del porcentaje de error más que de su magnitud.

Correlación entre los valores predichos y los observados. Se regresiona el cambio en la variable observada (y) contra el cambio en el valor predicho. El R^2 es usado como una medida de exactitud en la predicción.

Porcentaje de predicción de puntos de cambio (*turning points*). Esta medida es más apropiada cuando se está modelando una variable dicotómica, es decir, cuando la variable dependiente del modelo es una variable muda (*dummy*).[45] Dado que en estos modelos la predicción tiene que ser un uno o un cero, entonces, la medida de exactitud de la predicción se obtiene como el porcentaje de unos correctamente predichos menos el porcentaje de ceros incorrectamente predichos.[46]

Puntaje cuadrático. También aplica para modelos cuya variable dependiente es una *dummy*, e incluso es más popular que la medida anteriormente expuesta. Se calcula como (1-p^2) donde p es la probabilidad predicha para la ocurrencia del evento.

Eficiencia condicional. Esta medida es útil para comparar las predicciones de modelos alternativos. Una predicción A es condicionalmente eficiente relativa a la predicción B si B no posee información útil más allá de la contenida en A. Se regresiona la variable a predecir contra A y B y se testea la nula de que el coeficiente asociado a B es cero.

Finalmente, otra manera de evaluar el poder predictor de un modelo es regresionar el cambio en el valor observado sobre el cambio en el valor predicho más una

[45] Una variable *dummy* toma el valor uno (1) si se cumple una determinada condición y cero (0) si no.

[46] Estos modelos son muy útiles para modelar decisiones, como por ejemplo, viajar en avión, trabajar, ir al colegio, estudiar en la universidad, comprar una acción, ir al estadio, etc.

constante. Si la constante no es estadísticamente distinta de cero, y la pendiente no es estadísticamente distinta de uno, entonces, se dice que el predictor es bueno.

3.9. Algunas consideraciones

Existe un relativo consenso de que la mejor predicción podría ser un promedio ponderado de un conjunto de predictores, cada uno generado a partir de un modelo diferente. Así, si un investigador tuviera tres modelos alternativos para realizar una predicción, entonces, la mejor predicción que podría hacer sería la siguiente:

$$\hat{y} = \phi_1 \hat{y}^A + \phi_2 \hat{y}^B + \phi_3 \hat{y}^C$$

donde los ponderadores ϕ_i suman uno, pero pueden tener magnitudes distintas, no siendo necesariamente igual a un tercio cada uno.

Es necesario reconocer que en general a los econometristas no les va bien haciendo predicciones. Sin embargo, también se debe señalar que una mala predicción es mucho mejor que nada. Simon (1994) mencionó lo siguiente a este respecto: si bien en el corto plazo las predicciones económicas son malas, en el largo plazo son muy buenas ya que las leyes económicas tienden a imponerse en períodos largos de tiempo.

4. Violación de supuestos

El objetivo de este capítulo consiste en levantar los supuestos sobre los cuales se ha construido el modelo clásico de regresión lineal. De esta manera, se discutirán las consecuencias para el estimador de mínimos cuadrados ordinarios, y qué alternativas de solución existen.

4.1. Linealidad de los parámetros: el método de máxima verosimilitud (MV)

El primer supuesto sobre el cual se construyó el modelo clásico de regresión lineal fue el de linealidad. En particular se asumió que el modelo era lineal en los parámetros. Esto quiere decir que la relación existente entre los parámetros del modelo (no entre la variable dependiente y las variables explicativas) es lineal. En este apartado se verá qué alternativas de estimación existen cuando la relación entre los parámetros no es lineal. Para ello se deberá abandonar el método de estimación de MCO para presentar uno nuevo: el método de máxima verosimilitud. En términos generales, es posible señalar que existen dos tipos de modelos no lineales. Los modelos linealizables, que son aquellos que se pueden linealizar a través de una transformación simple. Por otro lado, están los modelos no linealizables, los cuales no pueden linealizarse mediante una operación directa. Para estos últimos se requiere utilizar métodos numéricos para estimar los parámetros poblacionales. También pueden estimarse por medio del método de mínimos cuadrados no lineales. Por ejemplo, considere la siguiente función de producción CES (elasticidad de sustitución constante):

$$y = \gamma[\delta K^{-\theta} + (1-\delta)L^{-\theta}]^{\frac{-\phi}{\theta}} e^{u}$$

Esta función no es lineal en los parámetros, y más aún no es linealizable a través de una operación matemática simple. Aplicando logaritmo se llega a lo siguiente:

$$ln(y) = ln(\gamma) - \frac{\phi}{\theta} ln[\delta K^{-\theta} + (1-\delta)L^{-\theta}] + u$$

Una alternativa de estimación es aplicar el método de mínimos cuadrados no lineales. Otra alternativa consiste en estimar mediante MV. En cambio, hay modelos que sí son linealizables. Por ejemplo, considere la siguiente función de producción:

$$y = \beta_1 X^{\beta_2} e^{u}$$

donde u es un shock estocástico distribuido idéntica e independientemente con media cero y varianza constante. Al aplicar logaritmo este modelo, que no es lineal en los parámetros, queda de la siguiente forma:

$$ln(y) = ln(\beta_1) + \beta_2 ln(X) + u$$

Este modelo es lineal en los parámetros y puede estimarse mediante MCO. La clave es que los parámetros se relacionen linealmente entre sí, pero la relación funcional entre la variable dependiente y las variables independientes no tiene que ser necesariamente lineal. De esta forma, los siguientes modelos son lineales en los parámetros:

$$ln(y) = \beta_1 + \beta_2 ln(X) + \beta_3 ln(Z) + u$$

$$ln(y) = \beta_1 + \beta_2 X + u$$

$$y = \beta_1 + \beta_2 X + u$$

$$y = \beta_1 + \beta_2 X + \beta_3 X^2 + u$$

$$y = \beta_1 + \beta_2 ln(L) + \beta_3 ln(K) + \beta_4 ln(K) ln(L) + \beta_5 ln(L)^2 + \beta_6 ln(K)^2 + u$$

La cuestión fundamental que viene a continuación se refiere a qué hacer en caso de que se detecte una no linealidad que no pueda ser capturada mediante una forma funcional más flexible. En ese caso, es conveniente analizar un método de estimación alternativo: el método de máxima verosimilitud.

Hasta el momento se ha adoptado el criterio de estimación que consiste en escoger los valores de los parámetros (β,σ^2) de modo de minimizar la suma de los residuos al cuadrado. Sin embargo, existe otra estrategia de estimación que, a diferencia de MCO, depende fuertemente en los supuestos de la distribución del término de error.[47] Este método se conoce como estimación por máxima verosimilitud (MV).

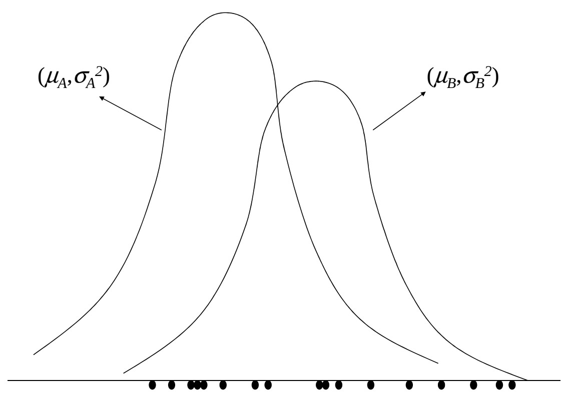
La estimación por máxima verosimilitud establece que los datos muestrales observados son más probables de provenir del mundo real caracterizado por un conjunto específico de parámetros (β,σ^2) que de otro cualquier otro conjunto de parámetros.

[47] Recuerde que sólo se realizó el supuesto de normalidad del término de error hacer inferencia.

En otras palabras, el estimador MV maximiza la probabilidad de ocurrencia de la muestra observada.

La figura 13 presenta el razonamiento anterior. Claramente, μ_B y σ_B^2 son los parámetros que maximizan la probabilidad de observar la muestra obtenida, y en consecuencia, el par B es preferido al par A.[48] Como se puede apreciar con los parámetros (β_B,σ_B^2) se abarca una mayor cantidad de observaciones muestales, que en este caso están representadas por cada uno de los puntos.

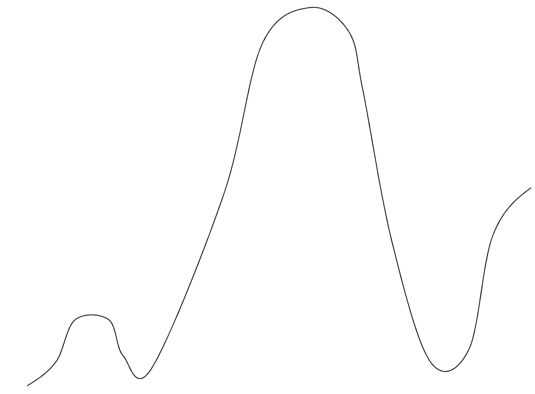
Figura 13: Estimación por máxima verosimilitud



Sin embargo, esta técnica es muy demandante en términos computacionales, sobre todo si la función de verosimilitud cuenta con más de un punto máximo (que por cierto no es el caso de la figura anterior). El algoritmo de estimación debe ser capaz de buscar a lo largo de toda la función de verosimilitud de manera de asegurar, con cierto grado de confiabilidad, que el máximo encontrado no es sólo local sino también global. En la práctica, muchas veces (en el caso de modelos altamente no lineales) estas funciones se maximizan a través de métodos numéricos mediante la utilización de algoritmos complejos de búsqueda. Uno de los algoritmos más populares es el método de Newton-Raphson. Estos algoritmos son muy útiles cuando el investigador se enfrenta a un problema de optimización cuya función objetivo (función de verosimilitud) tiene una forma como la que se presenta en la figura 14.

Figura 14: Ejemplo de una función a maximizar

[48] La principal desventaja de esta metodología radica en el hecho de que a priori se debe asumir una distribución específica para el término de error. Usualmente se escoge una distribución normal, pero ¿es sensato este supuesto? Si se piensa que el error consiste en una suma de shocks aleatorios, entonces, por el teorema central del límite se puede decir que se distribuirán de manera normal.



En este caso, además de verificar las condiciones de segundo orden para determinar si efectivamente se está en presencia de un punto máximo, se deberán realizar distintas búsquedas con el objetivo de encontrar un punto máximo global, y no local. Las técnicas iterativas de maximización, en general tienen la siguiente forma:

$$\theta^{**} = \theta^{*} + \lambda d(\theta^{*})$$

donde θ^{**} es el *update* (actualización) de la estimación de θ, θ^{*} es la estimación previa que se tiene, λ es un escalar positivo denominado *step-lenght* y $d(\theta^{*})$ es un vector direccional de dimensión k. Básicamente la distinción que existe entre los distintos métodos de búsqueda se da en $d(\theta^{*})$.

En vista del progreso que ha habido en términos computacionales, este método de estimación ha experimentado un rápido desarrollo durante los últimos años, ya que su principal característica (ventaja), a diferencia de la estimación tradicional de MCO, es que permite estimar modelos no lineales. Otra razón de su creciente popularidad radica en el hecho de que este estimador presenta muy buenas propiedades asintóticas. A continuación se hace un desarrollo formal de este método de estimación.

Sea:

$$y' = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}$$

un vector de dimensión n con valores muestrales para y, los cuales son dependientes de un vector de dimensión k de parámetros desconocidos:

$$\theta' = \begin{pmatrix} \theta_1 & \theta_2 & \cdots & \theta_k \end{pmatrix}$$

Se define la función de densidad conjunta como $f(y;\theta)$, que corresponde a la probabilidad de observar el vector y, probabilidad que depende del vector de parámetros θ. A esta función de densidad conjunta se le conoce también como función de verosimilitud (L):[49]

$$L(\theta; y) = f(y; \theta) \tag{52}$$

[49] En estricto rigor, si y toma valores continuos, la probabilidad de observar dicha muestra es cero. Por lo tanto, es más correcto señalar que el estimador MV es aquel que maximiza la probabilidad de obtener la muestra observada más menos cierta holgura, Δy.

Al maximizar la función de verosimilitud con respecto a θ se obtendrá un valor $\hat{\theta}$, que será aquel que maximiza la probabilidad de obtener los valores muestrales que se han observado (y), por lo tanto, θ corresponde al estimador de máxima verosimilitud (MV) del vector de parámetros desconocidos θ. El estimador de MV puede expresarse de la siguiente manera:

$$\hat{\theta}_{MV} \in argmax\ L(\theta; y)$$

Se debe notar que si las observaciones muestrales son independientes entre sí, entonces:

$$L(\theta; y) = \Pi_{i=1}^{n} L_i(\theta, y_i)$$

Es decir, la probabilidad (verosimilitud) conjunta corresponde al producto de las probabilidades individuales (L_i). Por conveniencia matemática se trabaja con el logaritmo de la función de verosimilitud, que corresponde a una transformación monotónica creciente de la función original, y por tanto, no altera la solución final del problema. Así, se define el logaritmo de la función de verosimilitud (denominada *log-likelihood*) como:

$$l = ln(L)$$

Por lo tanto, la solución al problema puede plantearse alternativamente de la siguiente forma:

$$\hat{\theta}_{MV} \in argmax\ l(\theta; y)$$

Luego:

$$l(\theta, y) = ln(\Pi_{i=1}^{n} L_i(\theta, y_i))$$

$$l(\theta, y) = \sum_{i=1}^{n} ln(L_i(\theta, y_i))$$

$$l(\theta, y) = \sum_{i=1}^{n} l_i(\theta, y_i)$$

Derivando con respecto a θ se llega a lo siguiente:

$$\frac{\partial l}{\partial \theta} = \frac{1}{L}\frac{\partial L}{\partial \theta}$$

A las derivadas de l con respecto a θ (recuerde que se trata de un vector de parámetros, por lo que hay k derivadas) se les denomina *score*. Luego, el estimador de MV ($\hat{\theta}$) se obtiene igualando el score a cero, por lo tanto:

$$s(\theta; y) = \frac{\partial l}{\partial \theta} = 0$$

Como puede apreciarse, la principal desventaja de este método de estimación radica en el hecho de que se debe asumir una distribución específica para el término de error, ya que solo así puede conocerce la verosimilitud.

A continuación se presentan las propiedades estadísticas del estimador de máxima verosimilitud, las cuales se cumplen en su totalidad solo si el modelo está correctamente especificado.[50]

[50]No obstante, si la función de verosimilitud no está especificada de manera correcta, de todas

4.1.1. Propiedades del estimador de máxima verosimilitud

La principal ventaja del estimador MV son sus propiedades asintóticas. Esto es muy importante pues en muchas ocasiones las estimaciones de MCO no presentan buenas propiedades en muestras finitas, por lo que el método de estimación debe escogerse en función de su comportamiento asintótico.

Consistencia. Un estimador $\hat{\theta}$ es consistente si cumple lo siguiente:

$$plim(\hat{\theta}) = \theta$$

Suponga una variable aleatoria g que se distribuye normal e independiente (iid) con media μ y varianza σ^2. Luego, se desprende que:

$$E(\bar{g}_n) = \mu$$

y:

$$var(\bar{g}_n) = \frac{\sigma^2}{n}$$

donde $\bar{g}_n$ corresponde a la media muestral. Por lo tanto, y tal como se puede apreciar, $\bar{g}_n$ es un estimador insesgado sin imoportar cuál sea el tamaño de la muestra, y además, la varianza tiende a cero cuando n crece indefinidamente. Se desprende, entonces, que sin importar la distribución de $\bar{g}_n$, a medida que crece n ésta se encuentra cada vez más concentrada en torno a un vecindario de μ. Sea:

$$P(\mu - \epsilon < \bar{g}_n < \mu + \epsilon) = P(|\bar{g}_n - \mu| < \epsilon)$$

la probabilidad de que $\bar{g}_n$ se encuentre en un intervalo determinado.[51] Dado que $var(\bar{g}_n)$ decrece monotónicamente a medida que aumenta n, entonces, existe un número n^* y un δ $(0 < \delta < 1)$ tal que para cualquier $n > n^*$:

$$P(|\bar{g}_n - \mu| < \epsilon) > 1 - \delta$$

Si esto ocurre, se dice que la variable aleatoria $\bar{g}_n$ converge en probabilidad a la constante μ. Equivalentemente, se tiene lo siguiente:

$$\lim_{n \to \infty} P(|\bar{g}_n - \mu| < \epsilon) = 1$$

Esto significa que la probabilidad de que $\bar{g}_n$ permanezca en un intervalo arbitrariamente pequeño en torno a μ puede ser muy cercana a uno dependiendo de qué tan grande sea n. Esto puede escribirse resumidamente como:

$$plim(\bar{g}_n) = \mu$$

donde el término *plim* indica la probabilidad límite. Por lo tanto, la media muestral es un estimador consistente de μ, ya que converge en probabilidad a ésta.

formas se obtienen estimaciones que son consistentes, y son asintóticamente normales. Ahora bien, si el modelo está bien especificado entonces además se logra eficiencia asintótica.

[51] Este intervalo podría ser muy pequeño, dependiendo de la elección que se haga de ϵ.

Algunas propiedades de la probabilidad límite son las siguientes:

$$plim(a_n b_n) = plim(a_n) \cdot plim(b_n)$$

y:

$$plim\left(\frac{a_n}{b_n}\right) = \frac{plim(a_n)}{plim(b_n)}$$

El operador esperanza no cumple con estas propiedades, a menos que a_N y b_N sean estocásticamente independientes.

Ahora suponga que se tiene otro estimador m_n para μ tal que:

$$E(m_n) = \mu + \frac{c}{n}$$

donde c es una constante cualquiera. Como puede apreciarse, este estimador es sesgado en muestras finitas, ya que $E(m_n) \neq \mu$, sin embargo:

$$\lim_{n\to\infty} E(m_n) = \mu$$

Si además se tiene que la varianza de m_n se va a cero a medida que aumenta n, entonces, m_n es un estimador consistente de μ.[52]

Normalidad asintótica. La normalidad asintótica se refiere a la distribución límite del estimador de máxima verosimilitud, la que vendrá dada por:

$$\hat{\theta} \sim N(\theta, I^{-1}(\theta))$$

Es decir, la distribución asintótica de $\hat{\theta}$ es normal con media cero y varianza dada por la inversa de la matriz de información ($I^{-1}(\theta)$). Esta varianza corresponde a la cota inferior de Cramer Rao, y es la menor varianza que puede lograr un estimador insesgado. La matriz de información se define de la siguiente manera:

$$I(\theta) = E\left[\left(\frac{\partial l}{\partial \theta}\right)\left(\frac{\partial l}{\partial \theta}\right)'\right] = -E\left[\frac{\partial^2 l}{\partial \theta \partial \theta'}\right]$$

Por lo tanto, hay dos formas de calcular la matriz de información, sin embargo, ocurre que muchas veces es más sencillo obtenerla a través de la expresión que incorpora la segunda derivada de la función $ln(L)$.

Cuando θ corresponde a un vector de dimensión k, entonces, $\frac{\partial l}{\partial \theta}$ denota un vector columna con las k derivadas parciales, esto es:

$$\frac{\partial l}{\partial \theta} = \begin{pmatrix} \frac{\partial l}{\partial \theta_1} \\ \frac{\partial l}{\partial \theta_2} \\ \vdots \\ \frac{\partial l}{\partial \theta_k} \end{pmatrix}$$

[52]Este caso corresponde a un ejemplo de convergencia en media cuadrática, que ocurre cuando el límite del valor esperado del estimador es igual al parámetro poblacional, y el límite de la varianza del estimador es cero. Convergencia en media cuadrática es una condición suficiente para probar consistencia, y se utiliza a menudo para establecer una probabilidad límite.

Derivando nuevamente con respecto a cada elemento de θ, por ejemplo, para θ_1 se obtiene lo siguiente:

$$\frac{\partial[\partial l/\partial\theta]}{\partial\theta_1} = \begin{pmatrix} \frac{\partial^2 l}{\partial\theta_1^2} & \frac{\partial^2 l}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 l}{\partial\theta_1\partial\theta_k} \end{pmatrix}$$

donde la segunda derivada ha sido escrita como un vector fila. Realizando esto para cada uno de los elementos del vector θ se obtiene una matriz cuadrada de segundas derivadas, que en este caso tiene dimensión k:

$$\frac{\partial^2 l}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{\partial^2 l}{\partial\theta_1^2} & \frac{\partial^2 l}{\partial\theta_1\partial\theta_2} & \cdots & \frac{\partial^2 l}{\partial\theta_1\partial\theta_k} \\ \frac{\partial^2 l}{\partial\theta_2\partial\theta_1} & \frac{\partial^2 l}{\partial\theta_2^2} & \cdots & \frac{\partial^2 l}{\partial\theta_2\partial\theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial\theta_k\partial\theta_1} & \frac{\partial^2 l}{\partial\theta_k\partial\theta_2} & \cdots & \frac{\partial^2 l}{\partial\theta_k^2} \end{pmatrix}$$

A esta matriz se le conoce con el nombre de *Hessiano* (H)

Eficiencia asintótica. Como se señaló anteriormente, el estimador de MV alcanza la cota inferior de Cramer Rao, esto es, su varianza viene dada por el inverso de la matriz de información. Más formalmente, esta propiedad puede expresarse de la siguiente manera: si $\hat{\theta}$ (escalar) es el estimador de máxima verosimilitud del parámetro θ, entonces:

$$\sqrt{n}(\hat{\theta} - \theta) \to N(0, \sigma^2)$$

donde σ^2 denota una constante finita. La propiedad de eficiencia asintótica establece que el estimador de MV tiene la mínima varianza de entre todos los estimadores consistentes y asintóticamente normales.[53] El término varianza asintótica se refiere a la varianza de la distribución límite del estimador, es decir cuando el tamaño de la muestra tiende a infinito. De igual manera, puede decirse que la varianza asintótica de $\hat{\theta}$ es σ^2/n. Por otro lado, cuando θ es un vector de parámetros y θ es el estimador de MV, lo anterior se plantea de la siguiente forma:

$$\sqrt{n}(\hat{\theta} - \theta) \to N(0, V)$$

donde V es una matriz definida positiva. Si $\tilde{V}$ denota la matriz de varianza de cualquier otro estimador consistente y asintóticamente normal, entonces, la propiedad de eficiencia asintótica establece que $\tilde{V} - V$ es una matriz semidefinida positiva.

4.1.2. Máxima verosimilitud y el modelo de regresión lineal

A continuación se derivará el estimador de MV del modelo de regresión lineal, y tal como se podrá apreciar, cuando el modelo a estimar es lineal y los errores se distribuyen de manera normal, el estimador de MCO y de MV coinciden. En efecto, suponga que se desea estimar el siguiente modelo:

$$y_i = \beta_1 + \beta_2 X_{i2} + \cdots + \beta_k X_{ik} + u_i$$

[53] Note que esto es más general que lo establecido en el teorema de Gauss-Markov, pues ahí la comparación se restringía solo a los estimadores lineales e insesgados.

con $u \sim N(0, \sigma^2)$. La función de verosimilitud para los errores u_i con $i = 1, ..., n$ viene dada por:

$$f(u_1, u_2, ..., u_n; \sigma^2 I) = f(u_1) \cdot f(u_2) \cdots f(u_n) = \Pi_{i=1}^n f(u_i)$$

Luego, dado que el término de error se distribuye normal, la función de densidad viene dada por:

$$f(u_i; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{u_i^2}{2\sigma^2}\right)$$

Así, la función de verosimilitud (probabilidad conjunta) vendrá dada por:

$$f(u_1, u_2, ..., u_n; \sigma^2 I) = \Pi_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{u_i^2}{2\sigma^2}\right)$$

Desarrollando la expresión anterior:

$$f(u_1, u_2, ..., u_n; \sigma^2 I) = (2\pi\sigma^2)^{-n/2} exp\left(-\frac{\sum_{i=1}^n u_i^2}{2\sigma^2}\right)$$

Sin embargo, se sabe que:

$$\sum_{i=1}^n u_i^2 = u'u$$

Por lo tanto:

$$f(u_1, u_2, ..., u_n; \sigma^2 I) = (2\pi\sigma^2)^{-n/2} exp\left(-\frac{u'u}{2\sigma^2}\right)$$

Esta expresión corresponde a la densidad normal multivariada para u. Por otro lado, la densidad multivariada de Y condicional en X, que es la importa en este contexto pues se quiere estimar β, viene dada por:

$$f(y|X) = f(u) \left|\frac{\partial u}{\partial y}\right|$$

donde $|\partial u / \partial y|$ es el valor absoluto del determinante formado a partir de la matriz cuadrada de dimensión n de derivadas parciales de los elementos de u con respecto a los elementos de y.[54] En este caso, es directo notar que esta matriz corresponde simplemente a la matriz identidad, ya que u_i es solo función de y_i. Por lo tanto, la función de densidad de y condicional a X coincide con la función de densidad de u, ya que el determinante de la matriz identidad es igual a uno. Luego:

$$L = f(y, X; \sigma^2, \beta) = (2\pi\sigma^2)^{-n/2} exp\left(-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}\right)$$

[54] Corresponde a:

$$\frac{\partial u_i}{\partial y_j}$$

para todo i,j.

Aplicando logaritmo a la expresión anterior:

$$ln(L) = ln\left((2\pi\sigma^2)^{-n/2}exp\left(-\frac{(y-X\beta)'(y-X\beta)}{2\sigma^2}\right)\right)$$

El estimador MV es aquel que maximiza la siguiente expresión:

$$\max_{\beta,\sigma^2} ln(L) = \max_{\beta,\sigma^2} ln\left((2\pi\sigma^2)^{-n/2}exp\left(-\frac{(y-X\beta)'(y-X\beta)}{2\sigma^2}\right)\right)$$

Finalmente:

$$\max_{\beta,\sigma^2} ln(L) = \max_{\beta,\sigma^2}\left(-\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{(y-X\beta)'(y-X\beta)}{2\sigma^2}\right) \tag{53}$$

Luego, las condiciones de primer orden (CPO) vienen dadas por:

$$\frac{\partial ln(L)}{\partial\beta} = -\frac{1}{\hat{\sigma}^2}X'(Y-X\hat{\beta}) = 0$$

Despejando para $\hat{\beta}$:

$$\hat{\beta}_{MV} = (X'X)^{-1}X'Y$$

Ahora derivando (51) con respecto a σ^2:

$$\frac{\partial ln(L)}{\partial\sigma^2} = -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}(Y-X\hat{\beta})'(Y-X\hat{\beta}) = 0$$

Despejando para $\hat{\sigma}^2$:

$$\hat{\sigma}^2_{MV} = \frac{(Y-X\hat{\beta})(Y-X\hat{\beta})}{n} = \frac{\hat{u}'\hat{u}}{n}$$

Se debe notar que en este contexto (linealidad y normalidad en el término de error) el estimador de MV de β coincide con el estimador de MCO, y la estimación de la varianza del error viene dada por $\hat{\sigma}^2_{MV} = \hat{u}'\hat{u}/n$. Antertiormente se demostró que $E(\hat{u}'\hat{u}/(n-k)) = \sigma^2$, por lo tanto, $E(\hat{\sigma^2_{MV}}) = \sigma^2(n-k)/n$, así es que $\hat{\sigma}^2_{MV}$ es un estimador sesgado (pero consistente) de σ^2. Además, $\hat{\beta}$ es un estimador insesgado y consistente de β.

Por otro lado, las segundas derivadas de la función de verosimilitud (*log-likelihood*) vienen dadas por:

$$\frac{\partial^2 l}{\partial\beta\partial\beta'} = -\frac{X'X}{\sigma^2}$$

Por lo tanto:

$$-E\left(\frac{\partial^2 l}{\partial\beta\partial\beta'}\right) = \frac{X'X}{\sigma^2}$$

Por otro lado:

$$\frac{\partial^2 l}{\partial\beta\partial\sigma^2} = -\frac{1}{\hat{\sigma}^4}X'(Y-X\hat{\beta}) = -\frac{X'\hat{u}}{\hat{\sigma}^4}$$

Luego:

$$-E\left(\frac{\partial^2 l}{\partial\beta\partial\sigma^2}\right) = 0$$

Además:

$$\frac{\partial^2 l}{\partial(\sigma^2)^2} = \frac{n}{2\hat{\sigma}^4} - \frac{\hat{u}'\hat{u}}{\hat{\sigma}^6}$$

Por lo tanto:

$$-E\left(\frac{\partial^2 l}{\partial(\sigma^2)^2}\right) = \frac{n}{2\hat{\sigma}^4}$$

dado que $E(\hat{u}'\hat{u}) = n\hat{\sigma}^2$. Utilizando estos resultados, la matriz de información queda de la siguiente manera:

$$I(\theta) = -E\left(\frac{\partial^2 l}{\partial\theta\partial\theta'}\right) = \begin{pmatrix} \frac{1}{\hat{\sigma}^2}(X'X) & 0 \\ 0 & \frac{n}{2\hat{\sigma}^4} \end{pmatrix}$$

y su inversa es:[55]

$$I^{-1}(\theta) = \begin{pmatrix} \hat{\sigma}^2(X'X)^{-1} & 0 \\ 0 & \frac{2\hat{\sigma}^4}{n} \end{pmatrix}$$

Note que:

$$var(\hat{\beta}_{MV}) = \hat{\sigma}^2(X'X)^{-1}$$

y:

$$var(\hat{\sigma}^2_{MV}) = \frac{2\hat{\sigma}^4}{n}$$

Así, en términos más generales, cuando se lleva a cabo una estimación por MV, para encontrar la matriz de varianzas y covarianzas basta con encontrar los elementos que constituyen la matriz de información.

4.1.3. Aplicación

A modo de ejemplo, considere la siguiente función de densidad:

$$f(y_i|X_i) = \frac{\lambda e^{-\lambda y_i}(\lambda y_i)^{X_i}}{X_i!}$$

donde $X_i, y_i > 0$. Para encontrar el estimador de MV de λ, se debe maximizar la función de verosimilitud. El logaritmo de la función de verosimilitud viene dado por:

$$ln(L) = ln\left(\Pi_{i=1}^n \frac{\lambda e^{-\lambda y_i}(\lambda y_i)^{X_i}}{X_i!}\right)$$

[55]La inversa de una matriz diagonal se obtiene invirtiendo cada uno de los elementos de la diagonal de la matrix.

Luego:

$$ln(L) = \sum_{i=1}^{n} ln\left(\frac{\lambda e^{-\lambda y_i}(\lambda y_i)^{X_i}}{X_i!}\right)$$

$$ln(L) = \sum_{i=1}^{n}[ln(\lambda) - \lambda y_i + X_i(ln(\lambda) + ln(y_i)) - ln(X_i!)]$$

$$ln(L) = nln(\lambda) - \lambda\sum_{i=1}^{n} y_i + ln(\lambda)\sum_{i=1}^{n} X_i ln(y_i) - \sum_{i=1}^{n} ln(X_i!)$$

A continuación se deriva con respecto a λ:

$$\frac{\partial ln(L)}{\partial\lambda} = \frac{n}{\hat{\lambda}} - \sum_{i=1}^{n} y_i + \frac{\sum_{i=1}^{n} X_i}{\hat{\lambda}} = 0$$

Finalmente:

$$\hat{\lambda} = \frac{n + \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} y_i}$$

$$\hat{\lambda} = \frac{1 + \bar{X}}{\bar{y}}$$

Luego, dado una muestra, el estimador $\hat{\lambda}$ es aquel que maximiza la probabilidad de ocurrencia de dicha muestra.

4.1.4. Inferencia en el contexto de MV

Los test que se desarrollarán en este apartado se realizarán en el contexto de un conjunto de hipótesis lineales (aunque estos test también sirven para hipótesis no lineales) acerca del vector de parámetros poblacionales β, de la forma:

$$H_0 : R\beta = r$$

donde R es una matriz de qxk ($q < k$) de constantes conocidas y r es un vector de qx1 de constantes conocidas. Recuerde que q corresponde al número de hipótesis a testear.

Test de razón de verosimilitud (LR). Considerando los estimadores de MV desarrollados en la sección anterior es posible calcular el valor maximizado de la función de verosimilitud como $L(\hat{\beta}, \hat{\sigma^2})$. A este valor se le conoce como función de máxima verosimilitud no restringida (ya que no se han impuesto los valores de la hipótesis nula en su cálculo), y es posible expresarla como función de la suma de los errores al cuadrado. Sin embargo, el modelo también podría ser estimado de una manera restringida (es decir, imponiendo los valores establecidos en las hipótesis nulas). Por ejemplo, maximizar la función de versomilitud sujeto a la restricción $R\beta = r$. Suponga que los estimadores resultantes de esta maximización restringida son $\bar{\beta}$ y $\bar{\sigma}^2$. De la misma manera, es posible obtener la función de MV restringida como $L(\bar{\beta}, \bar{\sigma}^2)$. Se debe notar que el máximo restringido no puede exceder el máximo no restringido.[56] Luego, la razón de verosimilitud se define de la siguiente manera:

$$\lambda = \frac{L(\bar{\beta}, \bar{\sigma}^2)}{L(\hat{\beta}, \hat{\sigma}^2)} \tag{54}$$

Intuitivamente, se debería rechazar la hipótesis nula si λ fuese pequeño, ya que esto sería indicativo de que la función restringida se aleja mucho de su valor sin restringir. Para muestras grandes se define la siguiente expresión para el test LR:

$$LR = -2ln(\lambda) = 2[lnL(\hat{\beta}, \hat{\sigma}^2) - lnL(\bar{\beta}, \bar{\sigma}^2)] \sim \chi^2(q) \tag{55}$$

Es decir, se distribuye como una chi-cuadrado con q grados de libertad. Así, si el valor calculado excede el valor crítico de tabla, entonces, existe evidencia para rechazar la hipótesis nula.

El estimador de MV restringido se deriva maximizando la siguiente expresión:

$$l^* = l - \mu'(R\beta - r)$$

donde μ es el vector que contiene los multiplicadores de Lagrange, y es de dimensión qx1. El vector que maximiza esta expresión corresponde a $\bar{\beta}$ y $\bar{\sigma}^2$. Finalmente, si el shock estocástico se distribuye normal el test LR puede escribirse como:

$$LR = n[ln(\bar{u}'\bar{u}) - ln(\hat{u}'\hat{u})]$$

Nuevamente, si el valor calculado excede el valor crítico de tabla, entonces, existe evidencia para rechazar la hipótesis nula.

El test de Wald (W). Para este test sólo se debe resolver el modelo no restringido, lo que se constituye en su principal ventaja. El vector $(R\hat{\beta} - r)$ indicaría el grado al cual la estimación del modelo restringido ajusta la hipótesis nula. Si este vector fuese cercano a cero, entonces, existiría evidencia para no rechazar la hipótesis nula. En cambio, un valor grande para este vector sería un indicativo de la invalidez de la hipótesis nula. Dado que $\hat{\beta}$ se distribuye asintóticamente normal con media vectorial β y matriz de varianzas y covarianzas $I^{-1}(\beta)$ se tiene que, bajo H_0, $(R\hat{\beta} - r)$ tiene una distribución asintótica multivariada normal con media vectorial cero y matriz de varianzas y covarianzas $RI^{-1}(\beta)R'$, donde $I^{-1}(\beta) = \sigma^2(X'X)^{-1}$. Por lo tanto:

$$(R\hat{\beta} - r)'[RI^{-1}(\hat{\beta})R']^{-1}(R\hat{\beta} - r) \sim \chi^2(q)$$

donde q es el número de restricciones en R. La distribución asintótica se mantiene cuando σ^2 es reemplazado por un estimador consistente, como por ejemplo, $\hat{\sigma}^2$. Cuando el shock estocástico se distribuye normal, al reemplazar el inverso de la matriz de información, el test de Wald toma la siguiente forma:

$$W = \frac{(R\hat{\beta} - r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta} - r)}{\hat{\sigma}^2} \sim \chi^2(q) \tag{56}$$

[56] Un óptimo restringido puede ser a lo más tan bueno como un máximo no restringido, pero no mejor.

Es posible demostrar que en este caso el test se reduce a:

$$W = \frac{n(\bar{u}'\bar{u} - \hat{u}'\hat{u})}{\hat{u}'\hat{u}}$$

Si el valor calculado excede el valor crítico de tabla, entonces, existe evidencia para rechazar la hipótesis nula.[57]

El test de los multiplicadores de Lagrange (LM). Al test LM se le conoce también con el nombre de *test score* ya que se basa en el score, y su principal ventaja es que solo requiere calcular el modelo restringido. El *score* viene dado por:

$$s(\theta) = \frac{\partial ln(L)}{\partial \theta} = \frac{\partial l}{\partial \theta}$$

El estimador no restringido $\hat{\theta}$ se obtiene resolviendo la expresión anterior para $s(\hat{\theta}) = 0$. Si las restricciones son válidas, entonces, el máximo restringido, $l(\bar{\theta})$, debiera ser cercano al máximo no restringido, $l(\hat{\theta})$. Es posible demostrar que el score tiene media vectorial cero, y matriz de varianzas y covarianzas dada por la matriz de información $I(\theta)$. Por lo tanto, la forma cuadrática $s'(\theta)I^{-1}s(\theta)$ tendrá una distribución χ^2. Evaluar esta forma cuadrática en $\bar{\theta}$ provee un test para la hipótesis nula. Luego, bajo la hipótesis nula se cumple lo siguiente:

$$LM = s'(\bar{\theta})I^{-1}(\bar{\theta})s(\bar{\theta}) \sim \chi^2(q)55 \tag{57}$$

Como se puede apreciar, y en contraste con lo que ocurre con el test de Wald, sólo se requiere calcular el estimador restringido ($\bar{\beta}$).[58] En el caso del modelo de regresión lineal múltiple, el score es el siguiente:

$$s(\theta) = \begin{pmatrix} \frac{\partial l}{\partial \beta} \\ \frac{\partial l}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} X'u \\ -\frac{n}{2\sigma^2} + \frac{u'u}{2\sigma^4} \end{pmatrix}$$

Luego, para evaluar el score en $\bar{\theta}$ se reemplaza u por $\hat{u}_* = Y - X\bar{\beta}$ y σ^2 por $\bar{\sigma}^2 = \hat{u}_*'\hat{u}_*/n$. El vector $\bar{\beta}$ satisface $R\bar{\beta} = r$. Por lo tanto:

$$s(\bar{\theta}) = \begin{pmatrix} \frac{1}{\bar{\sigma}^2} X'\hat{u}_* \\ 0 \end{pmatrix}$$

Por otro lado, la matriz de información evaluada en $\bar{\theta}$ viene dada por:

$$I^{-1}(\bar{\theta}) = \begin{pmatrix} \bar{\sigma}^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\bar{\sigma}^4}{n} \end{pmatrix}$$

Reemplazando los términos correspondientes en la ecuación (55):

$$LM = \begin{pmatrix} \frac{1}{\bar{\sigma}^2}\hat{u}_*'X & 0 \end{pmatrix} \begin{pmatrix} \bar{\sigma}^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\bar{\sigma}^4}{n} \end{pmatrix} \begin{pmatrix} \frac{1}{\bar{\sigma}^2} X'\hat{u}_* \\ 0 \end{pmatrix}$$

[57] Es directo notar que el test W es igual a q veces el test F.

[58] En ocasiones es más fácil calcular este estimador que el estimador no restringido.

Desarrollando la expresión anterior:

$$LM = \frac{\hat{u}_*' X(X'X)^{-1} X' \hat{u}_*'}{\tilde{\sigma}^2} = \frac{n\hat{u}_*' X(X'X)^{-1} X' \hat{u}_*'}{\hat{u}_*' \hat{u}_*}$$

Finalmente:

$$LM = nR^2 \tag{58}$$

donde R^2 es el coeficiente de determinación de una regresión de $\hat{u}_*$ sobre X. De esta manera, el test LM puede implementarse en dos pasos. En primer lugar se resuelve el modelo restringido y se calcula $\hat{\beta}$ para así obtener $\hat{u}_*$. Luego se hace una regresión de $\hat{u}_*$ sobre X. El estadístico nR^2 se distribuye $\chi^2(q)$. Si el valor calculado excede el valor crítico de tabla $(\chi^2(q))$, entonces, existe evidencia para rechazar la hipótesis nula.

Por otro lado, si el shock estocástico se distribuye normal, el test LM se puede escribir como:

$$LM = \frac{n(\hat{u}_*' \hat{u}_* - \hat{u}' \hat{u})}{\hat{u}_*' \hat{u}_*}$$

Finalmente, los tres test vistos hasta ahora se relacionan de la siguiente forma:

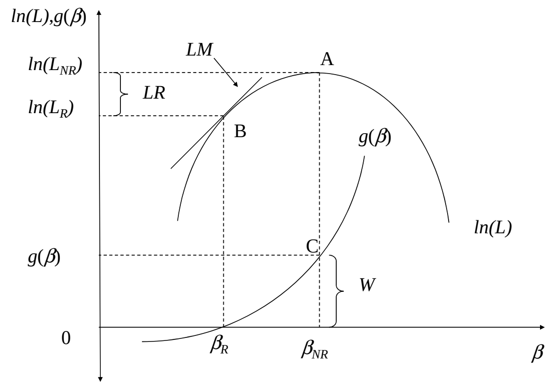
$$W \geq LR \geq LM \tag{59}$$

Sin embargo, los tres test son asintóticamente equivalentes, pero difieren en muestras pequeñas, lo que quiere decir que pueden generar resultados contradictorios entre sí.

4.1.5. Explicando los test LR, W y LM

El test F es posible de aplicar siempre que las restricciones sean lineales. Sin embargo, cuando se quiere testear restricciones no lineales, o cuando el modelo no sea lineal en los parámetros, o bien, cuando los errores no tengan una distribución normal, ya no es posible aplicar el test F y se debe recurrir a otros procedimientos, como por ejemplo, los test LR, W y LM. Como ya se ha señalado, estos test son asintóticamente equivalentes, y tienen una distribución χ^2 con grados de libertad igual al número de hipótesis a testar (q).

Figura 15: Explicando los test LR, W y LM



En la figura 15 se ha graficado la función $ln(L)$, y se muestra su relación con β. Note que en β_{NR} (es decir, el estimador no restringido) la función alcanza un valor máximo. Suponga que se está testeando la siguiente restricción $g(\beta) = 0$, la cual se satisface en el punto β_R, es decir, donde la restricción $g(\beta)$ corta a la abscisa.

En el contexto del test LR, cuando la restricción es cierta, el valor de la función de verosimilitud cuando se impone la restricción ($ln(L_R)$) no debiera ser significativamente menor que $ln(L_{NR})$, es decir, el valor máximo que alcanza la función sin imponer restricción alguna. Luego, el test LR evalúa si la diferencia entre $ln(L_R)$ y $ln(L_{NR})$ es estadísticamente distinta de cero. Si lo es, entonces, existe evidencia para rechazar la hipótesis nula. En efecto, recuerde que el test viene dado por:

$$LR = -2ln(\lambda) = 2[lnL(\hat{\beta}, \hat{\sigma}^2) - lnL(\bar{\beta}, \bar{\sigma}^2)] \sim \chi^2(q)$$

Por otro lado, en el contexto del test de Wald, si la restricción $g(\beta) = 0$ es verdadera, entonces, $g(\beta_{NR})$ (la estimación no restringida de β) no debiera ser estadísticamente distinta de cero. En consecuencia, el test W evalúa si β_{NR} viola la restricción por una cantidad significativa. En efecto:

$$(R\hat{\beta} - r)'[RI^{-1}(\hat{\beta})R']^{-1}(R\hat{\beta} - r) \sim \chi^2(q)$$

Finalmente, para el test LM, note que la función de verosimilitud se maximiza en el punto A, donde la pendiente de $ln(L)$ respecto a β es cero. Luego, si la restricción es verdadera, entonces, la pendiente de $ln(L)$ en el punto B no debiera ser estadísticamente distinta de cero. Así, el test LM testea si la pendiente de $ln(L)$, evaluada en la estimación restringida, es estadísticamente distinta de cero, y por ello es que utiliza el score, es decir, la pendiente de la función $ln(L)$.

$$LM = s'(\bar{\theta})I^{-1}(\bar{\theta})s(\bar{\theta}) \sim \chi^2(q)$$

4.1.6. Otros comentarios acerca de MV

Hasta el momento, en el contexto de la estimación por MV, se ha asumido que las observaciones muestrales son independientes entre sí, sin embargo, podría no ser el caso.[59] De hecho, las observaciones podrían no ser independientes, al menos, por dos razones:

1. Si el valor rezagado de la variable dependiente aparece como regresor ($y_t = \beta_1 + \beta_2 y_{t-1} + u_t$), lo cual es muy probable que ocurra en el contexto de las series de tiempo

2. Si los errores se encuentran autocorrelacionados

El primer punto solo podría darse en el contexto de series de tiempo, no en datos de corte transversal. El segundo punto si bien podría darse en el contexto de datos de corte transversal (autocorrelación espacial), es más probable que ocurra también en datos de series de tiempo. Ahora bien, si este es el escenario, hay dos cosas que se pueden hacer: se podría utilizar una función de distribución multivariada que incorpore las autocorrelaciones, o bien, pudiera realizarse una transformación. Considere la primera alternativa. Una función de distribución multivariada entrega la densidad del vector completo de u en lugar de sólo un elemento de éste (indica la probabilidad de obtener el conjunto completo de u). Por ejemplo, la función de densidad de una distribución normal multivariada (en términos matriciales) para u viene dada por:

$$f(u) = (2\pi\sigma^2)^{-n/2}|det(\Omega)|^{-1/2}e\left[\frac{1}{-2\sigma^2}u'\Omega^{-1}u\right]$$

donde $\sigma^2\Omega$ es la matriz de varianzas y covarianzas del vector u. Esta fórmula en sí se utiliza como la función de verosimilitud, por lo tanto, no es necesario multiplicar las densidades individuales, ya que aquí se cuenta con la distribución del vector completo de u. Note además que esta expresión corresponde a la densidad de u no de y, por lo que se debe ajustar por el término $|\partial u/\partial y|$, es decir, por el Jacobiano de la transformación de u a y.

La otra alternativa consiste, como se dijo, en realizar una transformación. Considere el siguiente modelo:

$$y_t = \beta_1 + \beta_2 X_t + \beta_3 Z_t + u_t$$

con $u_t = \rho u_{t-1} + e_t$ donde $e \sim N(0, \sigma_e^2)$. Dado que las observaciones de los u no son independientes, entonces, debe utilizarse la función de densidad de la distribución normal multivariada. En este contexto la matriz Ω va a estar en función de ρ, y σ_e^2. Sin embargo, dado que los valores e_t son independientes unos de otros, entonces, puede construirse la función de densidad conjunta de éstos multiplicando las densidades individuales de cada e_t. Dado que $u_t = \rho u_{t-1} + e_t$, entonces, $e_t = u_t - \rho u_{t-1}$. Por lo tanto:

$$e_t = (y_t - \rho y_{t-1}) - \beta_1(1-\rho) - \beta_2(X_t - \rho X_{t-1}) - \beta_3(Z_t - \rho Z_{t-1})$$

[59] Por ello, la función de densidad conjunta, la función de verosimilitud, se ha construido como la pitatoria (productoria) de las funciones de densidad individuales.

Luego, la densidad de y se calcula como como la densidad del vector e multiplicada por el Jacobiano de la transformación de e a y. Para este ejemplo, este método es más fácil de implementar que el primero puesto que aquel requiere calcular el determinante de Ω, lo que puede ser bastante complicado.

4.2. Forma funcional y especificación

Para poder estimar correctamente los parámetros poblacionales se requiere haber modelado correctamente la función de regresión poblacional. Por ello, la correcta especificación del modelo a estimar constituye un supuesto clave de la estimación por MCO. Así, el objetivo se esta sección consiste en analizar distintas formas funcionales que puden ser consideradas al momento de modelar la función de regresión poblacional.

Tal como se pudo apreciar en la secciones anteriores, la relación funcional entre la variable dependiente y las independientes puede ser altamente no lineal, el punto es cómo saber si es que la especificación escogida captura de buena forma la relación teórica verdadera, la cual lamentablemente es desconocida. Para ello existen varios test que permiten detectar no linealidades omitidas en el modelo estimado cuando la forma funcional es conocida. Por ejemplo, el test de Ramsey-Reset se implementa de la siguiente manera. Considere el siguiente modelo:

$$y_i = x_i'\beta + u$$

Para detectar alguna no linealidad no omitida, el test propone reestimar el modelo anterior pero incluyendo como variabls explicativas las potencias de la predicción de la variable dependiente del modelo original:

$$y_i = x_i'\beta + \alpha_2 \hat{y}_i^2 + \alpha_3 \hat{y}_i^3 + \cdots + \alpha_m \hat{y}_i^m + u$$

Si bien en este caso se incluyen potencias de grado m en este ejemplo, en la práctica basta con incluir potencias cuadradas y cúbicas. Luego, se hace un test de wald sobre los coeficientes de las variables incluidas, en particular, se testa la siguiente hipótesis nula: $\alpha_2 = \cdots = \alpha_m = 0$. Si se rechaza la hipótesis nula, entonces, existe evidencia de no linealidades omitidas.[60]

Los residuos recursivos son otra forma de detectar no linealidades. Se define al enésimo residuo recursivo como el error que se comete al predecir la enésima observación (y_n) utilizando los parámetros estimados con las primeras $n-1$ observaciones (regresión lineal). Si la verdadera función es no lineal, entonces, al ordenar los datos de acuerdo a la variable que (supuestamente) no es lineal, los errores podrían llegar a ser (irían aumentando o disminuyendo), o todos positivos, o bien, todos negativos. También se pueden implementar formas funcionales generales. Por ejemplo, algunas formas funcionales contienen a otras formas particulares, tales como la linealidad o la log linealidad, al asumir valores específicos para un parámetro. La idea sería

[60]Este test podría implementarse mediante un test F, sin embargo, para ello se requiere que el término de error se distribuya normal. En cambio, el test de Wald ($W = qF$) tiene la ventaja que siempre se distribuye Chi-cuadrado, sin importar la distribución del término de error.

testar aquellos valores. Por otro lado, los test no anidados se emplean básicamente para comparar modelos. Están también los test de cambio estructural, en donde se testa la existencia de regímenes en los datos (recuerde que una función no lineal puede ser aproximada por dos o más segmentos lineales). Un mecanismo alternativo para identificar no linealidades es dividir la muestra en grupos (de acuerdo al tamaño de la variable independiente que contendría la no linealidad), y comparar los parámetros estimados. Finalmente, el test de *Rainbow* (Utts, 1982), realiza estimaciones con datos tomados a partir de una cierta parte de la muestra, y luego, realiza predicciones para los puntos que no estaban incluidos en esta parte de la muestra. La idea es que las predicciones pertenezcan a la zona de confianza.

Una vez que se han detectado no linealidades omitidas en el modelo el desafío consiste en intentar capturarlas. Utilizar potencias de las variables independientes, o bien, intereacciones entre estas constituye un buen mecanismo de búsqueda de dichas no linealidaes.

En general, existen varias alternativas para llevar a cabo el modelamiento de la función de regresión poblacional. El uso de variables dummies y de la variable dependiente rezagada (en datos de series de tiempo) son muy útiles en la práctica, y por eso se pasan a considerar a continuación.

4.2.1. Variables dummies

El uso de variables dummies (mudas o dicotómicas) permite incorporar más flexibilidad al modelo. Estas variables sirven para capturar las características cualitativas de los datos. Por ejemplo, considere que está interesado en modelar la relación existente entre salarios y años de escolaridad. En este contexto el género de la persona puede ser también un factor relevante para explicar los salarios de los individuos, el punto es cómo incorporar esta característica en el modelo a estimar. La respuesta es mediante las variables dummies. En este caso, la variable dummy hombre (H) se definiría de la siguiente manera:

$$H_i = \begin{cases} 1 & \text{si el individuo es hombre} \\ 0 & \text{si no} \end{cases}$$

Luego, el modelo a estimar sería el siguiente:

$$y_i = \beta_1 + \beta_2 S_i + \beta_3 H_i + u_i$$

donde y es el logaritmo (natural) del salario, S son los años de escolaridad, y u es un shock estocástico bien comportado. Del mismo modo, podría haberse definido la vaiable dummy mujer como:

$$M_i = \begin{cases} 1 & \text{si el individuo es mujer} \\ 0 & \text{si no} \end{cases}$$

Por cierto, hay muchas otras características que podrían incorporarse en una ecuación de salarios como ésta.[61] Por ejemplo, pudieran incorporarse variables dum-

[61] A este modelo que relaciona el salario de las personas con los años de escolaridad y otras características se le conoce como ecuación de Mincer.

mies para la región en que vive el individuo, el sector económico en el que trabaja (industria, minería, construcción, agricultura, etc.), condición étnica, oficio que desempeña, entre otras.

Suponga ahora que el modelo es el siguiente:

$$y_i = \beta_1 M_i + \beta_2 H_i + u_i$$

En este contexto, note que no se ha incluido una constante. En este caso, el coeficiente asociado a la variable dummy representa el valor esperado de la variable dependiente para la categoría correspondiente; así β_1 representaría el salario promedio de las mujeres, mientras que β_2 el salario promedio de los hombres.[62] Suponga ahora que se deja fuera una categoría, y el modelo queda:

$$y_i = \beta_1 M_i + u_i$$

Si la estimación incluye un intercepto, la categoría omitida es la línea base o de comparación. Por lo tanto, el coeficiente asociado al resto de las categorías representa el grado en el cual éstas difieren de la categoría base. En este caso, β_1 representa el grado en el que difieren los salarios de las mujeres con respecto al de los hombres (categoría base). ¿Es posible incluir ambas categorías y un intercepto en el modelo econométrico? Por ejemplo, considere la siguiente formulación:

$$y_i = \beta_1 + \beta_2 M_i + \beta_3 H_i + u_i$$

En este caso, no es posible estimar por MCO el modelo anterior. El problema es que la matriz X sufre de colinealidad exacta en sus columnas, con lo que se viola el supuesto o confición de rango columna completo. Note que en este caso la matriz X toma la siguiente forma:

$$X = \begin{pmatrix} 1 & M_{11} & H_{12} \\ 1 & M_{21} & H_{22} \\ 1 & M_{31} & H_{32} \\ \vdots & \vdots & \vdots \\ 1 & M_{n1} & H_{n2} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 0 & 1 \end{pmatrix}$$

[62]Por ejemplo, en este contexto, el estimador MCO β_1 y β_2 vendría dado por:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1}X'y = \begin{pmatrix} M'M & M'H \\ H'M & H'H \end{pmatrix}^{-1} \begin{pmatrix} M'y \\ H'y \end{pmatrix}$$

donde M y H son los vectores columnas de las variables explicativas. Luego:

$$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \sum M_i^2 & \sum M_i H_i \\ \sum M_i H_i & \sum H_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \sum M_i y_i \\ \sum H_i y_i \end{pmatrix} = \begin{pmatrix} n_M & 0 \\ 0 & n_H \end{pmatrix}^{-1} \begin{pmatrix} \sum M_i y_i \\ \sum H_i y_i \end{pmatrix} = \begin{pmatrix} \bar{y}_M \\ \bar{y}_H \end{pmatrix}$$

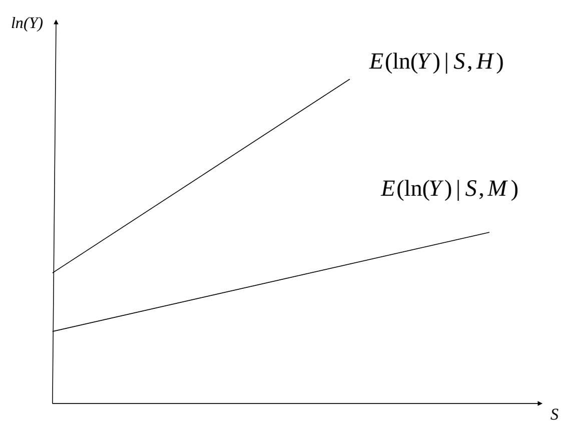
donde n_M y n_H corresponden al número de mujeres y de hombres que hay en la muestra, respectivamente. Además, la matriz de varianzas y covarianzas vienda dada por:

$$var(\hat{\beta}_1) = \sigma^2 (X'X)^{-1} = \sigma^2 \begin{pmatrix} n_M & 0 \\ 0 & n_H \end{pmatrix}^{-1} = \begin{pmatrix} \frac{\sigma^2}{n_M} & 0 \\ 0 & \frac{\sigma^2}{n_H} \end{pmatrix}$$

Dado que M y H son ceros o unos, note que la primera columna de la matriz X es una combinación lineal exacta de las columnas M y H (de hecho, es la suma). Luego, la matriz X es singular, lo que quiere decir que su determinante es cero,y por ende, no es invertible. Si no se puede invertir, entonces, no se puede estimar el vector de parámetros como $(X'X)^{-1}X'y$. Por lo tanto, cuando se tenga una variable que tenga g categorías, y se quiera colocar un intercepto en el modelo, solo se pueden incorporar $g-1$ variables dummies. Si se excluye el intercepto, en ese caso podrían incorporarse las g variables dummies, una para cada categoría. En el ejemplo anterior, $g = 2$, hombre y mujer, por ende, con intercepto solo se podría colocar una variable dummy.

Cabe señalar que existen dos tipos de variables dummies: variables dummies aditivas, y variables dummies multiplicativas. Ambos tipos son muy útiles al momento de capturar relaciones funcionales más flexibles. Volviendo al ejemplo de los salarios. Considere que se desea incorporar en la estimación dos caracteristicas de la relación que existe entre los salarios y los años de escolaridad: primero, que los hombres perciben en promedio un mayor salario que las mujeres, y segundo, que por cada año adicional de escolaridad los hombres obtienen un mayor retorno salarial. La figura 16 muestra esta situación.

Figura 16: Salarios y años de escolaridad



El modelo matemático que captura esta forma funcional es el siguiente:

$$y_i = \beta_1 + \beta_2 S_i + \beta_3 H_i + \beta_4(S_i H_i) + u_i$$

En este caso, la variable dummy H aparece de dos formas: por una parte es aditiva ($\beta_3 H_i$) pues cambia el intercepto de la ecuación y por otra es multiplicativa ($\beta_4(S_i H_i)$) pues altera la pendiente de la relación. De hecho, esto se ve más claramente al calcular el salario esperado de hombres y mujeres por separado. El salario esperado de los hombres viene dado por:

$$E(y|H_i = 1) = \beta_1 + \beta_2 S_i + \beta_3 + \beta_4 S_i = (\beta_1 + \beta_3) + (\beta_2 + \beta_4)S_i$$

Por otro lado, el salario esperado de las mujeres es:

$$E(y|H_i = 0) = \beta_1 + \beta_2 S_i$$

Es directo notar que tanto el intercepto $((\beta_1 + \beta_3))$ como la pendiente $((\beta_2 + \beta_4))$ de la ecuación para hombres es mayor que para las mujeres. La evidencia empírica disponible no solo para Chile sino que para el resto del mundo muestra que esto es efectivamente así. Esto ha dado origen a una vasta literatura económica que aborda la problemática de la discriminación salarial.

La tabla 3 presenta la estimación de la ecuación de Mincer pero incorporando una variable dummy que toma el valor 1 si la persona es mujer y cero en caso contrario. Dado que la variable dependiente está medida en logaritmos, la estimación muestra que las mujeres perciben en promedio un salario 15 % más bajo que el de los hombres. El coeficiente de esta variable es significativo al 1 %.

Tabla 3: Variables dummies
Dependent Variable:LWPH
Method: Least Squares
Sample: 12900
Included observations: 2900

Variable	**Coefficient**	**Std.Error**	**t-Statistic**	**Prob.**
C	5.414159	0.110796	48.86605	0.0000
ESC	0.096326	0.002887	33.36656	0.0000
EDAD	0.025327	0.005606	4.517529	0.0000
EDAD2	-0.000154	7.00E-05	-2.196855	0.0281
MUJER	-0.152533	0.022694	-6.721312	0.0000
R-squared	0.286383	Mean dependent var		7.112859
Adjusted R-squared	0.285397	S.D dependent var		0.649708
S.E of regression	0.549225	Akaike info criterion		1.641107
Sum squared resid	873.2725	Schwarz criterion		1.651404
Log likelihood	-2374.605	Hannan-Quinn criter.		1.644817
F-statistic	290.4489	Durbin-Watsons stat		2.064743
Prob (F-statistic)	0.000000			

Otro ejercicio interesante consiste en incorporar también una variable dummy multiplicativa con los años de escolaridad y determinar cuál es el retorno salarial para las mujeres por cada año de escolaridad. La tabla 4 muestra los resultados de la estimación con los datos de la encuesta CASEN 2009.

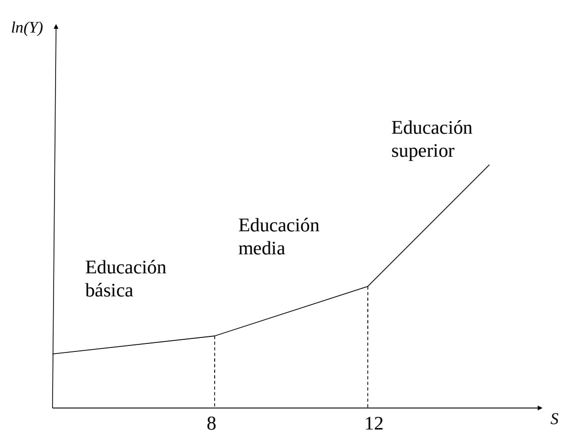
Tabla 4: Variables dummies multiplicativas
Dependent Variable:LWPH
Method: Least Squares
Sample: 12900
Included observations: 2900

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	5.449477	0.111853	48.72003	0.0000
ESC	0.092223	0.003424	26.93435	0.0000
EDAD	0.025840	0.005607	4.608253	0.0000
EDAD2	-0.000162	7.00E-05	-2.315308	0.0207
MUJER	-0.308727	0.073784	-4.184188	0.0000
ESCMUJER	0.013673	0.006146	2.224584	0.0262
R-squared	0.287601	Mean dependent var		7.112859
Adjusted R-squared	0.286370	S.D dependent var		0.649708
S.E of regression	0.548851	Akaike info criterion		1.640088
Sum squared resid	871.7817	Schwarz criterion		1.652445
Log likelihood	-2372.128	Hannan-Quinn criter.		1.644540
F-statistic	233.6658	Durbin-Watsons stat		2.064133
Prob (F-statistic)	0.000000			

Es interesante notar que el coeficiente asociado a la variable dummy multiplicativa es positivo, es decir, que por cada año adicional de escolaridad las mujeres obtienen en promedio un mayor retorno salarial que los hombres, el cual es estadísticamente significativo al 5 por ciento. No obstante lo anterior, note que la brecha salarial después de controlar por las variables de capital humano (escolaridad, y edad) se mantienen, tal como lo revela el coeficiente de la variable dummy aditiva.

Otra aplicación interesante con variables dummies se refiere a las regresiones tipo *spline*. Por ejemplo, suponga que se desea estimar el modelo planteado en la figura 17, que nuevamente hace referencia a la relación existente entre los salarios y los años de escolaridad. Un aspecto interesante que permite capturar la regresión spline es la existencia de un retorno o premio diferenciado a los años de escolaridad dependiendo del ciclo educativo al que haya llegado el individuo. Así, es esperable que el retorno salarial por cada año adicional de educación básica sea menor que el obtenido por cada año de educación media, y éste menor a su vez que cada año adicional de educación superior.

Figura 17: Regresión *Spline*



El modelo que captura este comportamiento puede expresarse de la siguiente manera:

$$y_i = \beta_1 + \beta_2 S_i + \beta_3 (S_i - 8) B_i + \beta_4 (S_i - 12) M_i + u_i$$

En este caso, las variables dummies se definen de la siguiente forma:

$$B_i = \begin{cases} 1 & \text{si el individuo tiene educación básica completa } (S_i \geq 8) \\ 0 & \text{si no} \end{cases}$$

$$M_i = \begin{cases} 1 & \text{si el individuo tiene educación media completa } (S_i \geq 12) \\ 0 & \text{si no} \end{cases}$$

Por lo tanto, ¿cuál es el salario esperado de un individuo que tiene educación básica incompleta? Para ello se debe obtener el valor esperado de la expresión anterior condicional a que $B_i = 0$ y $M_i = 0$:

$$E(y_i | B_i = 0, M_i = 0) = \beta_1 + \beta_2 S_i$$

Luego, el retorno a la escolaridad en este caso es:

$$\frac{\partial E(y_i | B_i = 0, M_i = 0)}{\partial S} = \beta_2$$

Por otro lado, el salario esperado para un individuo que tiene educación básica completa y media incompleta es:

$$E(y_i | B_i = 1, M_i = 0) = \beta_1 + \beta_2 S_i + \beta_3 (S_i - 8) B_i$$

En este caso, el retorno o premio a la escolaridad es:

$$\frac{\partial E(y_i | B_i = 1, M_i = 0)}{\partial S} = \beta_2 + \beta_3$$

Finalmente, el salario esperado para alguien que tiene educación media completa es:

$$E(y_i | B_i = 1, M_i = 1) = \beta_1 + \beta_2 S_i + \beta_3 (S_i - 8) B_i + \beta_4 (S_i - 12) M_i$$

El retorno a la escolaridad en este caso es:

$$\frac{\partial E(y_i | B_i = 1, M_i = 0)}{\partial S} = \beta_2 + \beta_3 + \beta_4$$

Como es posible apreciar, en esta ecuación se captura la relación entre salarios y años de escolaridad que aparece en la figura 17. La tabla 5 presenta los resultados de estimar esta ecuación de salarios utilizando los datos para Chile.

Tabla 5: Variables dummies y regresión spline
Dependent Variable:LWPH
Method: Least Squares
Sample: 12900
Included observations: 2900

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	5.908112	0.113916	51.86368	0.0000
ESC	0.040642	0.007669	5.299308	0.0000
MEDIA	0.024295	0.012596	1.928821	0.0539
SUPERIOR	0.110042	0.011359	9.687565	0.0207
EDAD	0.022588	0.005506	4.102538	0.0000
EDAD2	-0.000155	6.86E-05	-2.266476	0.0235
R-squared	0.324497	Mean dependent var		7.112859
Adjusted R-squared	0.323330	S.D dependent var		0.649708
S.E of regression	0.534449	Akaike info criterion		1.586907
Sum squared resid	826.6305	Schwarz criterion		1.599264
Log likelihood	-2295.015	Hannan-Quinn criter.		1.591359
F-statistic	278.0433	Durbin-Watsons stat		2.056429
Prob (F-statistic)	0.000000			

Los resultados de la estimación muestra un perfil convexo para los retornos a la escolaridad en Chile por nivel educacional. En efecto, para la educación básica el retorno salarial de cada año adicional de escolaridad es de 4 %, para la educación media es de 6.4 %, mientras que para la educación superior es de 17.4 %.

Algunos comentarios finales sobre el uso de variables dummies. En el contexto de series de tiempo, cuando por ejemplo, se está modelo el comportamiento del PIB o bien las ventas de una empresa, se puede incorporar una variable dummy para una observación específica, lo que permitiría capturar caídas o subidas (quiebres) en el comportamiento de la serie. Por otro lado, dado que el uso de variables dummies aumenta el número de regresores en el modelo, su uso exagerado podría generar artificialmente un elevado R^2. Finalmente, cabe señalar, tammbién en el contexto de series de tiempo el uso de variables dummies permite realizar ajustes por estacionalidad, es decir, sirven para desestacionalizar datos.[63]

4.2.2. Variable dependiente rezagada

Es posible que se quiera modelar el comportamiento de una variable en función de su propia historia, y en realidad esta es una estrategia bastante común en el contexto de series de tiempo. En efecto, ciertas variables presentan altos grados de persistencia en su comportamiento como por ejemplo, la cantidad de dinero, el PIB, el consumo, o la inversión. Es por ello que los modelos autorregresivos de orden p, AR(p), son muy populares. La variable dependiente tiene como variable explicativa a ella misma, pero rezagada un período, dos períodos, hasta p períodos. En otras palabras, la variable se explica por rezagos de ella misma. Es de orden p pues depende de los primeros p rezagos. Considere a modo de ejemplo, el siguiente

[63]En general, una serie puede descomponerse en cuatro elementos: (1) tendencia-ciclo: movimientos de largo plazo, (2) Estacionalidad: movimientos recurrentes en ciertos períodos del año, (3) Efecto calendario: impacto de la estructura del calendario sobre una serie, (4) Irregularidad: errores estadísticos o de eventos accidentales (u). Mediante el uso de variables dummies es posible extraer el componente de estacionalidad. Por cierto, también permitiría tomar en cuenta el denominado efecto calendario. Todo esto permitiría identificar la tendencia ciclo de la serie, es decir, los fundamentos subyacentes del comportamiento de largo plazo de la serie.

modelo AR(1):[64]

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

con $|\beta_1 < 1|$.[65] En esta ecuación, al término β_1 se le conoce como *drift*. Por ende, cuando se hace mención a un modelo AR(1) con *drift* es sencillamente para indicar que el modelo posee una constante. Un modelo AR(2) tendría la siguiente formulación:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + u_t$$

Considere ahora el siguiente modelo AR(1) sin *drift*, en donde las variables se encuentran en desvíos con respecto a su media:

$$y_t = \beta y_{t-1} + u_t$$

con $|\beta < 1|$. Luego, el estimador MCO para β vendrá dado por:[66]

$$\hat{\beta} = \frac{\sum_{t=2}^{T} y_t y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}$$

Desarrollando la expresión:

$$\hat{\beta} = \frac{\sum_{t=2}^{T} (\beta y_{t-1} + u_t) y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2} = \beta + \frac{\sum_{t=2}^{T} y_{t-1} u_t}{\sum_{t=2}^{T} y_{t-1}^2}$$

Aplicando esperanza a la expresión anterior:

$$\hat{\beta} = \beta + E\left(\frac{\sum_{t=2}^{T} y_{t-1} u_t}{\sum_{t=2}^{T} y_{t-1}^2}\right)$$

Por lo tanto, para que $\hat{\beta}$ sea insesgado se requiere que y_s y u_t sean independientes para todo t y s. Es posible calcular las covarianzas entre ambos términos para evaluar la veracidad de la afirmación anterior. En efecto, tomando la ecuación genérica para y_t es posible notar que:

$$y_1 = \beta y_0 + u_1$$

Por otro lado, para y_2:

$$y_2 = \beta y_1 + u_2 = \beta(\beta y_0 + u_1) + u_2 = \beta^2 y_0 + \beta u_1 + u_2$$

Recursivamente, se llega a la siguiente expresión para y_t:

$$y_t = \beta^t y_0 + \beta^{t-1} u_1 + \beta^{t-2} u_2 + \cdots + \beta u_{t-1} + u_t$$

[64]Es directo notar que en este contexto los regresores son estocásticos.

[65]Cuando $|\beta_1 < 1|$ se dice que la serie es estacionaria, y este es un requisito fundamental para que el modelo pueda estimarse por MCO. Cuando se cumple esta condición, el modelo tiene media y varianza. Cuando $|\beta_1 = 1|$ se dice que la serie tiene raíz unitaria, y por ende, el modelo no tiene media y varianza. De hecho, tanto la media como la varianza aumentan sin cota en el tiempo.

[66]Note que en la fórmula la sumatoria parte de $t = 2$ pues se pierde una observación al colocar un rezago de la variable dependiente al lado derecho. En realidad, se perderán tantas observaciones como rezagos de y se coloquen al lado derecho de la ecuación.

A continuación se puede calcular la covarianza entre y_t y u_t:

$$E(y_t u_t) = E[(\beta^t y_0 + \beta^{t-1} u_1 + \beta^{t-2} u_2 + \cdots + \beta u_{t-1} + u_t) u_t] = E(u_t^2) = \sigma^2$$

De la misma manera, se puede calcular la covarianza entre y_t y u_{t-1}:

$$E(y_t u_{t-1}) = E[(\beta^t y_0 + \beta^{t-1} u_1 + \beta^{t-2} u_2 + \cdots + \beta u_{t-1} + u_t) u_{t-1}] = \beta E(u_{t-1}^2) = \beta\sigma^2$$

Y para la covarianza entre y_t y u_{t-2} se tiene:

$$E(y_t u_{t-2}) = E[(\beta^t y_0 + \beta^{t-1} u_1 + \beta^{t-2} u_2 + \cdots + \beta u_{t-1} + u_t) u_{t-2}] = \beta^2 E(u_{t-2}^2) = \beta^2\sigma^2$$

Finalmente, note que:

$$E(y_t u_{t+s}) = 0$$

para todo $s \geq 1$. Un resultado interesante es que si bien $\hat{\beta}$ es sesgado, es consistente. De hecho, se puede demostrar que para este caso particular el sesgo viene dado por:

$$sesgo(\hat{\beta}) \approx -\frac{2\beta}{T}$$

es decir, el sesgo tiende a cero a medida que crece el tamaño de la muestra.[67] Un aspecto interesante que surge en el contexto de los modelos AR es la elección del número de rezagos. Ciertamente este no es un tema solo de los modelos AR sino más bien surge en todo proceso de modelamiento. Por ejemplo, en un modelo de salarios el econometrista podría estar indeciso respecto de si incluir la variable años de experiencia laboral como un regresor adicional. O bien, si se está intentando explicar el PIB de la economía, uno podría preguntarse cuántos rezagos se debieran incluir; ¿debiera ser un modelo AR(1) o AR(2)? Los criterios de selección de modelos han sido desarrollados precisamente con ese objetivo, seleccionar en base a criterios estadísticos cuál es la mejor formulación para explicar el comportamiento de la variable dependiente. Estos criterios de selección que se presentarán a continuación se pueden aplicar siempre que los modelos que se están comparando (es decir, de entre los cuales se quiere escoger uno) sean anidados, y además, el tamaño de la muestra de los distintos modelos debe ser el mismo.[68] A continuación se presentarán tres criterios de información ampliamente utilizados en el análisis econométrico, el criterio de Akaike (AIC), el criterio de Schwarz (BIC) y el criterio de Hannan-Quinn (HQC):

$$AIC = -\frac{2l}{T} + 2\left(\frac{k}{T}\right)$$

[67] Si el modelo tiene *drift* se puede demostrar que el sesgo viene dado por:

$$sesgo(\hat{\beta}) \approx -\frac{1}{T}(1 + 3\beta)$$

[68] Por lo tanto, hay que tener cuidado cuando se compara un modelo AR(1) con un modelo AR(2), pues este último realizará la estimación con una observación menos que el modelo AR(1). ¿Qué hacer entonces? Se debe quitar la primera observación al modelo AR(1) antes de llevar a cabo la estimación. Así se estarán comparando modelos comparables, pues ambos han sido estimados utilizando la misma información.

$$BIC = -\frac{2l}{T} + ln(T)\left(\frac{k}{T}\right)$$
$$HQC = -\frac{2l}{T} + 2ln(ln(T))\left(\frac{k}{T}\right)$$

donde T es el tamaño de la muestra con que se hace la estimación, l es la función log-likelihood, y k es el número de regresores incluidos en el modelo.[69] Dado que se debe escoger aquel modelo que presente el menor valor para el criterio de información, la inclusión de regresores adicionales sufre una penalización de orden 2 en el caso de AIC, de $ln(T)$ en el caso de BIC y de $2ln(ln(T))$ en el caso de HQC. El criterio de BIC es el que imputa el mayor grado de penalización, y por ende, tiende a seleccionar modelos más parsimoniosos. Por otro lado, el criterio de HQC impone una penalización menor que BIC pero mayor que AIC. Dado que el cálculo de los criterios de información requiere calcular la log-likelihood, cuando la estimación se haga por MCO se debe tener una estimación de ésta. No obstante, recuerde que si se estima por MCO y se asume que el error se distribuye de manera normal, la log-likelihood viene dada por:

$$l = -\frac{T}{2}ln(2\pi) - \frac{T}{2}ln(\hat{\sigma}^2) - \frac{\hat{u}'\hat{u}}{2\hat{\sigma}^2}$$

Sin embargo, el estimador sesgado de la varianza del error viene dado por:

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T}$$

Por lo tanto:

$$T = \frac{\hat{u}'\hat{u}}{\hat{\sigma}^2}$$

Así, reemplazando en la fórmula de la log-likelihood:

$$l = -\frac{T}{2}ln(2\pi) - \frac{T}{2}ln(\hat{\sigma}^2) - \frac{T}{2}$$

Luego:

$$-\frac{2l}{T} = ln(2\pi) + ln(\hat{\sigma}^2) + 1 \approx ln(\hat{\sigma}^2) = ln\left(\frac{\hat{u}'\hat{u}}{T}\right)$$

Reemplazando en las expresiones de los criterios de información, se llega a:

$$AIC \approx ln\left(\frac{\hat{u}'\hat{u}}{T}\right) + 2\left(\frac{k}{T}\right)$$

$$BIC \approx ln\left(\frac{\hat{u}'\hat{u}}{T}\right) + ln(T)\left(\frac{k}{T}\right)$$

$$HQC \approx ln\left(\frac{\hat{u}'\hat{u}}{T}\right) + 2ln(ln(T))\left(\frac{k}{T}\right)$$

Por lo tanto, para calcular los criterios de información cuando la estimación se ha llevado a cabo por MCO se utilizan los errores de MCO.

[69] Cabe señalar que los criterios de Schwarz y Hannan-Quinn son consistentes.

4.2.3. Elección de regresores: omisión de variable relevante e inclusión de variable irrelevante

En este apartado se analizarán dos problemas que son frecuentes en el análisis empírico al momento de escoger el conjunto de regresores que serán parte del modelo: omisión de variable relevante (OVR) e inclusión de variable irrelevante (IVI). Para analizar las consecuencias que tienen cada uno de estos problemas, considere el siguiente modelo planteado en términos matriciales:

$$y = X_1\beta_1 + X_2\beta_2 + u$$

donde y es un vector columna de dimensión n, X_1 es una matriz de dimensión nxk_1, β_1 es un vector de dimensión k_1, X_2 es una matriz de dimensión nxk_2, β_2 es un vector de dimensión k_2, y u es un vector de dimensión n. También se tiene que $k_1 + k_2 = k$, es decir, el modelo $y = X\beta + u$ se ha expresado de una forma distinta, para separar el vector de parámetros β en dos vectores β_1 y β_2. En efecto, note que:

$$X = \begin{pmatrix} X_1 & X_2 \end{pmatrix}$$

y:

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Por ende:

$$y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u$$

Recuerde que las ecuaciones normales de MCO vienen dadas por:

$$X'X\hat{\beta} = X'Y$$

Reemplazando por las nuevas expresiones para X y β:

$$\begin{pmatrix} X_1 & X_2 \end{pmatrix}' \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1 & X_2 \end{pmatrix}' y$$

Desarrollando la expresión anterior:

$$\begin{pmatrix} X_1' \\ X_2' \end{pmatrix} \begin{pmatrix} X_1 & X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1' \\ X_2' \end{pmatrix} y$$

Multiplicando los vectores:

$$\begin{pmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} X_1'y \\ X_2'y \end{pmatrix}$$

De esta forma, se llega al siguiente sistema de ecuaciones:

$$\begin{aligned} X_1'X_1\hat{\beta}_1 + X_1'X_2\hat{\beta}_2 &= X_1'y \\ X_2'X_1\hat{\beta}_1 + X_2'X_2\hat{\beta}_2 &= X_2'y \end{aligned}$$

De la segunda expresión se puede obtener:

$$\hat{\beta}_2 = (X_2'X_2)^{-1}X_2'(y - X_1\hat{\beta}_1)$$

Reemplazando ahora en la primera expresión:

$$X_1'X_1\hat{\beta}_1 + X_1'X_2(X_2'X_2)^{-1}X_2'(y - X_1\hat{\beta}_1) = X_1'y$$

Luego:

$$X_1'X_1\hat{\beta}_1 + X_1'P_2(y - X_1\hat{\beta}_1) = X_1'y$$

donde P_2 es la matriz de proyección. Luego de algunas manipulaciones:

$$X_1'(I - P_2)X_1\hat{\beta}_1 = X_1'(I - P_2)y$$

Luego:

$$X_1'M_2X_1\hat{\beta}_1 = X_1'M_2y$$

donde M_2 es la matriz de residuos.[70] Finalmente:

$$\hat{\beta}_1 = (X_1'M_2X_1)^{-1}X_1'M_2y$$

Repitiendo el procedimiento, se puede encontrar la siguiente expresión para el estimador MCO de β_2:

$$\hat{\beta}_2 = (X_2'M_1X_2)^{-1}X_2'M_1y$$

Es fácil demostrar que ambos estimadores son insesgados, y por ende, consistentes. Falta aún derivar la matriz de varianzas y covarianzas. Para ello note que:

$$\hat{\beta}_1 = (X_1'M_2X_1)^{-1}X_1'M_2y = (X_1'M_2X_1)^{-1}X_1'M_2(X_1\beta_1 + X_2\beta_2 + u)$$

Multiplicando los términos:

$$\hat{\beta}_1 = \beta_1 + (X_1'M_2X_1)^{-1}X_1'M_2X_2\beta_2 + (X_1'M_2X_1)^{-1}X_1'M_2u$$

Así:

$$\hat{\beta}_1 - \beta_1 = (X_1'M_2X_1)^{-1}X_1'M_2X_2\beta_2 + (X_1'M_2X_1)^{-1}X_1'M_2u$$

Con esta expresión es posible derivar la matriz de varianzas y covarianzas para $\hat{\beta}_1$:

$$var(\hat{\beta}_1) = E[(\hat{\beta}_1 - E(\hat{\beta}_1))(\hat{\beta}_1 - E(\hat{\beta}_1))']$$

Reemplazando y simplificando algunos términos:

$$var(\hat{\beta}_1) = E[((X_1'M_2X_1)^{-1}X_1'M_2u)((X_1'M_2X_1)^{-1}X_1'M_2u)']$$

Desarrollando la expresión e introduciendo el operador esperanza:

$$var(\hat{\beta}_1) = (X_1'M_2X_1)^{-1}X_1'M_2uE(uu')M_2X_1(X_1'M_2X_1)^{-1}$$

[70]Recuerde que las matrices de proyección y residuos son simétricas e idempotentes.

Finalmente:
$$var(\hat{\beta}_1) = \sigma^2(X_1'M_2X_1)^{-1}$$

De manera similar se obtiene la matriz de varianzas y covaianzas para $\hat{\beta}_2$:

$$var(\hat{\beta}_2) = \sigma^2(X_2'M_1X_2)^{-1}$$

Con estos resultados es posible evaluar los efectos que tiene la omisión de vaiables relevantes y la inclusión de variables irrelevantes. Considere primero el caso de la OVR. Suponga que el modelo verdadero viene dado por:

$$MV : y = X_1\beta_1 + X_2\beta_2 + u$$

Sin embargo, por alguna razón (descuido o bien falta de información) el econometrista estima el siguiente modelo (omitiendo k_2 variables relevantes):

$$ME : y = X_1\beta_1 + u$$

¿Qué consecuencias tiene esto para la estimación de β_1? El estimador MCO para β_1 vendrá dado por:
$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y$$

Luego, reemplazando y por su verdadera expresión:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + u)$$

Multiplicando términos:

$$\hat{\beta}_1 = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'u$$

Aplicando esperanza:

$$E(\hat{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'E(u)$$

Dado que el shock estocástico tiene media vectorial cero se llega a:

$$E(\hat{\beta}_1) = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2$$

Por lo tanto, el estimador MCO de β_1 es sesgado, y además inconsistente. El sesgo viene dado por:

$$E(\hat{\beta}_1) - \beta_1 = (X_1'X_1)^{-1}X_1'X_2\beta_2$$

Es decir, el sesgo depende de dos factores: (i) de la relación existente entre las variables omitidas y las variables incluidas en el modelo ($(X_1'X_1)^{-1}X_1'X_2$), y (ii) del efecto que tienen las variables omitidas sobre la variable dependiente (β_2).

¿Qué se puede decir de la matriz de varianzas y covarianzas? La varianza del modelo estimado viene dada por:

$$var(\hat{\beta}_1|X_1) = \sigma^2(X_1'X_1)^{-1}$$

Por otro lado, la matriz de varianzas y covarianzas del modelo verdadero:

$$var(\hat{\beta}_1|X_1, X_2) = \sigma^2(X_1'M_2X_1)^{-1}$$

Es posible mostrar que:[71]

$$var(\hat{\beta}_1|X_1, X_2) = \sigma^2(X_1'M_2X_1)^{-1} > \sigma^2(X_1'X_1)^{-1} = var(\hat{\beta}_1|X_1)$$

Por lo tanto, la omisión de variable relevante provoca sesgo en la estimación por MCO y además, la matriz de varianza y covarianza es más eficiente. Con respecto al sesgo que provoca la OVR, un ejemplo sencillo puede aclarar más la problemática. En efecto, considere el siguiente modelo para los salarios de las personas (y):

$$y_i = \beta_1 + \beta_2S_i + \beta_3H_i + u_i$$

donde S son los años de escolaridad, y H es la habilidad del individuo. Como siempre, u es un shock bien comportado con media cero y varianza homocedástica. El modelo plantea que los salarios dependen positivamente (β_2,$\beta_3 > 0$) de la escolaridad y de la habilidad de los individuos. Sin embargo, dado que el econometrista no observa la habilidad de los individuos, entonces, estimar el siguiente modelo:

$$y_i = \beta_1 + \beta_2S_i + u_i$$

Luego:

$$\hat{\beta}_2 = \frac{\sum s_iy_i}{\sum s_i^2}$$

donde la variable en minúsculas indica que se encuentra en desvíos con respecto a su media. Reeemplando y por su verdadera expresión:

$$\hat{\beta}_2 = \frac{\sum s_i(\beta_2s_i + \beta_3h_i + u_i)}{\sum s_i^2}$$

Por lo tanto:

$$\hat{\beta}_2 = \beta_2 + \beta_3\frac{\sum s_ih_i}{\sum s_i^2} + \frac{\sum s_iu_i}{\sum s_i^2}$$

Aplicando esperanza:

$$E(\hat{\beta}_2) = \beta_2 + \beta_3\frac{\sum s_ih_i}{\sum s_i^2}$$

Por lo tanto, el estimador MCO para β_2 es sesgado, y además inconsistente.[72]

[71] Esto se demuestra calculando la diferencia entre $var(\hat{\beta}_1|X_1, X_2)$ y $var(\hat{\beta}_1|X_1)$ y analizando si esta matriz es semi-definida positiva, es decir, se debe mostrar que los valores propios son positivos.

[72] En efecto, considere la expresión anterior:

$$\hat{\beta}_2 = \beta_2 + \beta_3\frac{(1/n)\sum s_ih_i}{(1/n)\sum s_i^2} + \frac{(1/n)\sum s_iu_i}{(1/n)\sum s_i^2}$$

donde se ha multiplicado tanto el numerador como el denominador por $1/n$. Luego, aplicando la

A continuación se analiza las consecuencias que tiene para la estimación por MCO la inclusión de variables irrelevantes. Suponga que el modelo verdadero viene dado por:

$$y = X_1\beta_1 + u$$

sin embargo, el econometrista incluye un conjunto de variables (k_2) irrelevantes en la estimación, es decir, estima el siguiente modelo:

$$y = X_1\beta_1 + X_2\beta_2 + u$$

El estimador MCO de β_1 viene dado por:

$$\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 y$$

Reemplazando y por su verdadera expresión:

$$\hat{\beta}_1 = (X_1' M_2 X_1)^{-1} X_1' M_2 (X_1\beta_1 + u) = \beta_1 + (X_1' M_2 X_1)^{-1} X_1' M_2 + u$$

Aplicando esperanza:

$$E(\hat{\beta}_1) = \beta_1$$

Por otro lado, la matriz de varianzas y covarianzas del modelo estimado viene dada por:

$$var(\hat{\beta}_1 | X_1, X_2) = \sigma^2 (X_1' M_2 X_1)^{-1}$$

y la matriz de varianzas y covarianzas del modelo verdadero viene dada por:

$$var(\hat{\beta}_1 | X_1) = \sigma^2 (X_1' X_1)^{-1}$$

Y como se sabe que $var(\hat{\beta}_1 | X_1) < var(\hat{\beta}_1 | X_1, X_2)$, entonces, las consecuencias que tiene la IVI es que el estimador MCO sigue siendo insesgado (y consistente), pero se pierde eficiencia.[73]

---

probabilidad límite ($plim$):

$$plim(\hat{\beta}_2) = \beta_2 + \beta_3 \frac{plim(1/n)\sum s_i h_i}{plim(1/n)\sum s_i^2} + \frac{plim(1/n)\sum s_i u_i}{plim(1/n)\sum s_i^2}$$

Las expresiones convergen a las varianzas y covarianzas poblacionales, por lo tanto:

$$plim(\hat{\beta}_2) = \beta_2 + \beta_3 \frac{cov(S,H)}{var(S)} + \frac{cov(S,u)}{var(S)}$$

Es decir:

$$plim(\hat{\beta}_2) = \beta_2 + \beta_3 \frac{cov(S,H)}{var(S)}$$

En este caso, dado que es esperable que $cov(S,H) > 0$ y dado que siempre $var(S) > 0$, entonces, el sesgo asintótico será positivo,y por ende, $plim(\hat{\beta}_2) > \beta_2$, es decir, el estimador MCO estará sesgado hacia arriba.

[73]Dado que la IVI aumenta las varianzas, note que los test t van a ser menores, y por ende, aumentará la probabilidad de que el test t de significancia individual se encuentre en la zona de no rechazo de la hipótesis nula. En otras palabras, note que el incremento en el número de variables irrelevantes puede hacer que las variables independientes no sean estadísticamente significativas, no porque no lo sean desde un punto de vista económico, sino más bien por un aumento en la varianza de los coeficientes estimados.

4.3. Perturbaciones no esféricas

Hasta ahora se ha asumido que los errores (o shocks estocásticos) son homocedásticos y no están autocorrelacionados, es decir, se ha asumido que:

$$E(u_i^2) = \sigma^2$$

para todo $i = 1, 2, \ldots, n$, y que:

$$E(u_i u_j) = 0$$

para todo $i \neq j$. Cuando ambos supuestos se cumplen se dice que las perturbaciones son esféricas. Cuando esto ocurre note que la matriz de varianzas y covarianzas de los errores toma la siguiente forma:

$$E(uu') = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix} = \sigma^2 I_n$$

El objetivo ahora consiste en levantar estos supuestos y analizar las consecuencias que tiene para la estimación por MCO la violación de éstos. En términos más formales, se asumirá lo siguiente:

$$E(u_i^2) = \sigma_i^2$$

y:

$$E(u_i u_j) = \sigma_{ij}$$

Por lo tanto, la matriz de varianzas y covarianzas de los shocks vendrá dada por:

$$E(uu') = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{n1} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{n2} \\ \vdots & \vdots & \ddots & 0 \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix} = \sigma^2 \Omega$$

donde Ω es una matriz cuadrada de dimensión nxn. Cuando la matriz de varianza y covarianza de los errores toma esta forma se dice que las perturbaciones son no esféricas. En vista de esta nueva estructura a continuación se analizarán las propiedades del estimador MCO. Para ello, considere la forma matricial del modelo de regresión lineal múltiple:

$$Y = X\beta + u$$

En este contexto se sabe que el estimador MCO de β viene dado por:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Reemplazando y:

$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u) = \beta + (X'X)^{-1}X'u$$

Aplicando esperanza:

$$E(\hat{\beta}) = \beta + (X'X)^{-1}X'E(u) = \beta$$

dado que el término de error tiene media vectorial cero. Por lo tanto, en presencia de perturbaciones no esféricas, el estimador MCO sigue siendo insesgado, y por ende, consistente. Falta por ver qué ocurre con la varianza. Para ello note que:

$$\hat{\beta} - \beta = (X'X)^{-1}X'u$$

Además, se sabe que:

$$var(\hat{\beta}) = E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)']$$

dado que el estimador MCO es insesgado. Reemplazando:

$$var(\hat{\beta}) = E[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'] = E((X'X)^{-1}X'uu'X(X'X)^{-1})$$

Introduciendo el operador esperanza:

$$var(\hat{\beta}) = (X'X)^{-1}X'E(uu')X(X'X)^{-1}$$

Luego:

$$var(\hat{\beta}) = (X'X)^{-1}X'\sigma^2\Omega X(X'X)^{-1}$$

Finalmente:

$$var(\hat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Note que esta expresión para la matriz de varianza y covarianza es muy distinta de aquella que se tenía cuando las perturbaciones eran esféricas ($\sigma^2(X'X)^{-1}$). De hecho, note que solo cuando $\Omega = I_n$ la expresión para la matriz de varianzas y covarianzas se reduce a:

$$var(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

En consecuencia, en presencia de perturbaciones no esféricas el estimador MCO sigue siendo insesgado, pero ya no es MELI, es decir, ya no es el estimador más eficiente. En este contexto, el estimador MELI será el estimador de mínimos cuadrados generalizados (MCG). Antes de derivar el estimador de MCG, se debe explicar en qué consiste la descomposición espectral de una matriz. A cualquier matriz que sea definida positiva, simétrica y finita, como por ejemplo, la matriz Ω, se le puede aplicar la descomposición espectral de la siguiente forma:

$$\Omega = C\Lambda C'$$

donde la matriz C está formada por los vectores propios de Ω, y Λ es una matriz diagonal cuyos elementos son los valores propios de Ω. Por otro lado, se debe tener presente que siempre se cumple que $C'C = CC' = I$, por lo tanto, $C' = C^{-1}$. Ahora se va a definir a la matriz P' como:

$$P' = C\Lambda^{-1/2}$$

por lo tanto:
$$P'P = C\Lambda^{-1/2}\Lambda^{-1/2}C' = C\Lambda^{-1}C' = \Omega^{-1}$$

Estos resultados se utilizarán a continuación para presentar el estimador de MCG. En efecto, considere el modelo de regresión lineal en su forma matricial:

$$y = X\beta + u$$

Premultiplicando el modelo por la matriz P se llega a:

$$Py = PX\beta + Pu$$

Luego:
$$y^* = X^*\beta + u^*$$

¿Qué propiedades exhibe este nuevo término de error u^*? Note que:

$$E(u^*u^{*\prime}) = E(Puu'P') = PE(uu')P' = P\sigma^2\Omega P' = \sigma^2 P\Omega P'$$

Reemplazando las matrices P, P' y Ω:

$$E(u^*u^{*\prime}) = \sigma^2(C\Lambda^{-1/2})'(C\Lambda C')(C\Lambda^{-1/2})$$

Luego:
$$E(u^*u^{*\prime}) = \sigma^2\Lambda^{-1/2}C'C\Lambda C'C\Lambda^{-1/2}$$

Simplificando términos:
$$E(u^*u^{*\prime}) = \sigma^2 I$$

Es decir, las perturbaciones del modelo transformado (u^*) son esféricas. Así, se puede aplicar la fórmula del estimador de MCO a este modelo transformado, lo que corresponde al estimador de MCG:

$$\hat{\beta} = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*$$

Reemplazando los términos:

$$\hat{\beta} = ((PX)'PX)^{-1}(PX)'Py = (X'P'PX)^{-1}X'P'Py$$

Dado que $P'P = \Omega^{-1}$, la ecuación queda:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

Esta expresión corresponde al estimador de MCG y este estimador, a diferencia de MCO, es MELI (BLUE). Es directo demostrar que este estimador es insesgado, y por ende, consistente:

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}(X\beta + u) = \beta + (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u$$

Aplicando esperanza:

$$E(\hat{\beta}) = \beta + (X'\Omega^{-1}X)^{-1}X'E(u) = \beta$$

dado que $E(u) = 0$. Para derivar la matriz de varianzas y covarianzas se debe notar que:

$$\hat{\beta} - \beta = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}u$$

Luego:

$$var(\hat{\beta}) = E[((X'\Omega^{-1}X)^{-1}X'u)((X'\Omega^{-1}X)^{-1}X'u)']$$

Desarrollando la expresión anterior:

$$var(\hat{\beta}) = E[(X'\Omega^{-1}X)^{-1}X'uu'X(X'\Omega^{-1}X)^{-1}]$$

Introduciendo el operador esperanza:

$$var(\hat{\beta}) = (X'\Omega^{-1}X)^{-1}X'E(uu)'X(X'\Omega^{-1}X)^{-1}$$

Reemplazando $E(uu')$:

$$var(\hat{\beta}) = \sigma^2(X'\Omega^{-1}X)^{-1}X'\Omega'X(X'\Omega^{-1}X)^{-1}$$

Simplificando términos:

$$var(\hat{\beta}) = \sigma^2(X'\Omega^{-1}X)^{-1}$$

Dado que σ^2 es desconocido, se debe estimar utilizando los errores de MCO del modelo transformado:

$$\tilde{\sigma}^2 = \frac{u^{*\prime}u^*}{n-k} = \frac{(y^* - X^*\hat{\beta}_{MCG})'(y^* - X^*\hat{\beta}_{MCG})}{n-k}$$

Luego:

$$\tilde{\sigma}^2 = \frac{(Y - X\hat{\beta}_{MCG})'P'P(Y - X\hat{\beta}_{MCG})}{n-k} = \frac{(Y - X\hat{\beta}_{MCG})'\Omega^{-1}(Y - X\hat{\beta}_{MCG})}{n-k}$$

Para hacer la inferencia estadística se procede de manera habitual pero teniendo presente que la varianza del error debe estimarse de esta manera.

Tal como se ha visto, todo el procedimiento que se ha aplicado hasta ahora depende de un factor fundamental: que la matriz Ω sea conocida. Si ese no es el caso, entonces, se debe estimar la matriz Ω, y aplicar el método conocido como mínimos cuadrados factibles (MCF). De esta forma, conociendo la forma de la matriz Ω se podría implementar el estimador de MCF como:

$$\hat{\beta} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y$$

Cabe señalar que el estimador MCF no es MELI, sin embargo posee buenas propiedades asintóticas. Además, no es lineal.

4.3.1. Heterocedasticidad

Con el objetivo de analizar más detenidamente las consecuencias que tiene la existencia de las perturbaciones no esféricas, se estudiará el problema en dos partes. En primer lugar, analizaremos el problema de la heterocedasticidad (varianzas distintas), y asumiremos que los errores no están autocorrelacionados, es decir, asumiremos que:

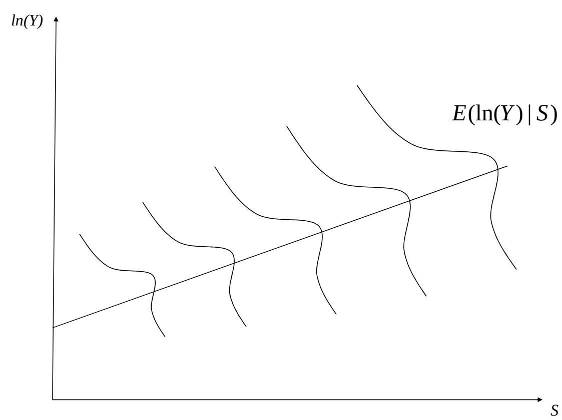
$$E(u_i^2) = \sigma_i^2$$

para todo $i = 1, 2, \ldots, n$. Y:

$$E(u_i u_j) = 0$$

para $i \neq j$. La figura 18 presenta graficamente en qué consiste el fenómeno denominado heterocedasticidad. Básicamente, lo que ocurre es que la varianza de cada observación es distinta, y puede existir un patrón para ello. Por ejemplo, como se puede apreciar en la figura 18, se observa que la varianza del término de error aumenta con el valor de la variable independiente.

Figura 18: Heterocedasticidad



En este caso, la figura plantea la relación (teórica) existente entre los salarios y los años de escolaridad. Como se puede apreciar, a medida que las personas exhiben mayores años de escolaridad aumenta la dispersión de los salarios en torno a la función de regresión poblacional ($\sigma^2 = f(S)$). ¿Por qué podría darse un patrón de estas características? Basicamente se puede dar puesto que mientras mayor sea la escolaridad de las personas factores como la habilidad, el poder de negociación, o la red de contactos, pueden constituir un factor clave para acceder a salarios más altos. En cambio, mientras menores sean los años de escolaridad, las personas pueden estar sujetas a un conjunto más acotado de salarios, encontrándose todos estos muy cercas a la función de regresión poblacional, sin posibilidades ciertas de diferenciarse mucho unos de otros. Ciertamente, dependiendo del modelo en cuestión se pueden encontrar distintos patrones para la forma de la varianza del término de error.

Tal como se estableció anterior, en presencia de heterocedasticidad, el estimador de MCO es insesgado pero ineficiente. Si se conoce la matriz Ω se aplica MCG, y si se desconoce Ω pero se sabe su forma, podría estimarse aplicando luego el estimador de MCF. Pero, ¿qué hacer en caso de que la matriz Ω sea completamente desconocida. En ese caso, White (1980) propuso el siguiente método para estimar la matriz de varianzas y covarianzas. Se sabe que:

$$var(\hat{\beta}) = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Reordenando términos:

$$var(\hat{\beta}) = n(X'X)^{-1}\left(\frac{1}{n}\sigma^2 X'\Omega X\right)(X'X)^{-1}$$

White propuso estimar el término que se encuentra entre paréntesis de la siguiente manera:

$$\Sigma = \frac{1}{n}\sigma^2 X'\Omega X = \frac{1}{n}\sum \sigma_i^2 x_i x_i'$$

donde x_i es un vector columna de dimensión kx1, que contiene la información de la i-ésima observación. Luego, la estimación de la matriz Σ es:

$$\hat{\Sigma} = \frac{1}{n}\sum \hat{u}_i^2 x_i x_i'$$

White demostró que la matriz $\hat{\Sigma}$ converge en probabilidad a Σ. Por lo tanto, la matriz de varianzas y covarianzas en presencia de heterocedasticidad se calcula como:

$$var(\hat{\beta}) = n(X'X)^{-1}\hat{\Sigma}(X'X)^{-1}$$

La tabla 6 presenta los resultados de estimar por Eviews nuevamente la ecuación de Mincer básica (salarios en contra de escolaridad, edad y edad al cuadrado) pero utilizando la corrección de White para estimar la matriz de varianzas y covarianzas de los parámetros del modelo.

Tabla 6: Corrección de White Dependent Variable:LWPH
Method: Least Squares
Sample: 12900
Included observations: 2900
White heteroskedasticity-consistent standard errors and covariance

Variable	**Coefficient**	**Std.Error**	**t-Statistic**	**Prob.**
C	5.908112	0.113916	51.86368	0.0000
ESC	0.040642	0.007669	5.299308	0.0000
EDAD	0.024295	0.012596	1.928821	0.0539
EDAD2	0.110042	0.011359	9.687565	0.0207
R-squared	0.275247	Mean dependent var		7.112859
Adjusted R-squared	0.274496	S.D dependent var		0.649708
S.E of regression	0.553399	Akaike info criterion		1.655902
Sum squared resid	886.8997	Schwarz criterion		1.664140
Log likelihood	-2397.057	Hannan-Quinn criter.		1.658870
F-statistic	366.6141	Durbin-Watsons stat		2.059148
Prob (F-statistic)	0.000000			

Note que los coeficientes estimados coinciden con los presentados en la tabla 2, sin embargo, dado que ahora la matriz de varianzas y covarianzas se estima tomando en cuenta la presencia de heterocedasticidad, los errores estándar, el test t y los p-value son distintos. Esto es así pues los coeficientes de estiman por MCO de la manera usual, sion embargo, se utiliza la fórmula correcta para encontrar las varianzas. Lo que podría ocurrir, por lo tanto, es que al tomar en cuenta la corrección de White algunos coeficientes dejen de ser estadísticamente significativos, o bien, que pasen a serlo.

Se han desarrollado varios test para detectar la presencia de heterocedasticidad, problema que es bastante común en datos de corte transversal. A continuación presentaremos solo los más utilizados.

Test de White. La hipótesis nula de este test viene dada por $H_0 : \sigma_i^2 = \sigma^2$, es decir, que existe homocedasticidad. El test se implementa de la siguiente manera: se estima el modelo original por MCO con lo cual se obtienen los residuos de MCO ($\hat{u}$). Luego, se corre una regresión auxiliar de los errores de MCO al cuadrado ($\hat{u}^2$) en contra de la variables explicativas del modelo original, sus interacciones y sus potencias al cuadrado. El estadístico nR^2 se distribuye Chi-cuadrado con q grados de libertad, donde el R^2 es el coeficiente de determinación de la regresión auxiliar, y q es el número de regresores incluidos en la regresión auxiliar. Si el valor nR^2 es mayor que el Chi-cuadrado crítico de tabla, se rechaza la hipótesis nula, y existe evidencia a favor de la presencia de errores heterocedásticos.

Test de Goldfeld y Quandt. Para implementar este test se debe saber con qué variable observable se relaciona la heterocedasticidad. Por ejemplo, suponga que $\sigma^2 = f(Z)$. Luego, se deben ordenar las observaciones según Z, y se omiten las p observaciones centrales de la muestra. Por lo tanto, se cuenta ahora con dos muestras o grupos de observaciones, y por ende, se estima el modelo con cada una de dichas muestras. La hipótesis nula es de homocedasticidad, y se calcula el siguiente estadístico (que se distribuye F con m grados de libertad en el numerador y en el denominador):[74]

$$\frac{\hat{u}_2'\hat{u}_2}{\hat{u}_1'\hat{u}_1}$$

Si el valor calculado para este estadístico es mayor que el F crítico, entonces, se rechaza la hipótesis nula.[75]

Test de Breusch-Pagan. Para este test se asume que la heterocedasticidad se relaciona con un conjunto de variables explicativas de la siguiente forma:

$$\sigma_i^2 = \alpha_0 + \alpha_1 Z_1 + \alpha_2 Z_2 + \cdots \alpha_p Z_p + \zeta_i$$

[74] El valor m se calcula como:

$$m = \frac{n-p}{2} - k$$

[75] Note que se ha asumido que $\sigma_1^2 < \sigma_2^2$

Luego, la hipótesis nula de homocedasticidad viene dada por: $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$. Para implementar este test primero se estima el modelo original por MCO y se obtiene los residuos de MCO al cuadrado ($\hat{u}^2$). Luego, se construyen los residuos estandarizados de la siguiente forma:

$$\hat{e}^2 = \frac{\hat{u}^2}{\hat{\sigma}^2}$$

A continuación se estima el siguiente modelo:

$$\hat{e}_i^2 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \cdots \gamma_p Z_p + \eta_i$$

Se obtiene la SEC (suma explicada de los cuadrados) y se calcula el siguiente estadístico:

$$\frac{SEC}{2}$$

que se distribuye Chi-cuadrado con p grados de libertad. Si este estadístico es mayor que el Chi-cuadrado crítico existe evidencia para rechazar la hipótesis nula y los errores serían heterocedásticos.

Test de Glesjer. Para implementar este test también se debe estimar el modelo original por MCO y calcular los errores al cuadrado ($\hat{u}^2$). Luego, se estiman varios modelos entre los errores al cuadrado y distintas combinaciones de los regresores; el que exista o no heterocedasticidad dependerá si los coeficientes estimados de estas regresiones auxiliares son o no estadísticamente significativos.

4.3.2. Autocorrelación

A continuación analizaremos el problema de la autocorrelación de los errores, y para ello se asumirá que los errores son homocedásticos. Por lo tanto, en este contexto el comportamiento de las perturbaciones es el siguiente:

$$E(u_i^2) = \sigma^2$$

para todo $i = 1, 2, \ldots, n$. Y:

$$E(u_i u_j) = \sigma_{ij}$$

para $i \neq j$. El problema de la autocorrelación es muy común en datos de series de tiempo, y casi inexistente en datos de corte transversal. Este fenómeno se da cuando el shock en el período t afecta el comportamiento del shock en el período $t+1$. Esto genera persistencia en el shock estocástico, la cual puede capturarse modelando el error a través de modelos autoregresivos.

En este contexto, la matriz de varianzas y covarianzas de los shocks estocásticos tendrá la siguiente forma:

$$E(uu') = \sigma^2 \Omega$$

Como ya se discutió anteriormente, si la matriz Ω es conocida, entonces s estima por MCG. Por otro lado, si dicha matriz es desconocida pero se conoce su forma, entonces, se puede implementar el método de MCF. Suponga que se conoce perfectamente la ecuación que describe el comportamiento de las perturbaciones, por ejemplo, asuma que:

$$y_t = x_t'\beta + u_t$$

con:

$$u_t = \rho u_{t-1} + e_t$$

Es decir, el término de error (u_t) es un proceso AR(1), y e_t es un término de error bien comportado (independiente e idénticamente distribuido) con media cero y varianza σ^2. Dada que la ecuación para el shock estocástico se asume conocida, es posible calcular las varianzas y covarianzas. En efecto, note que:

$$var(u_t) = \rho^2 var(u_{t-1}) + var(e_t) + 2cov(u_{t-1}, e_t)$$

Llamando σ_u^2 a la varianza incondicional de u_t, y dado que $cov(u_{t-1}, e_t)$ es cero (e_t es iid), entonces:

$$\sigma_u^2 = \rho^2\sigma_u^2 + \sigma^2$$

Luego, la varianza del término de error es:

$$\sigma_u^2 = \frac{\sigma^2}{1-\rho^2}$$

Con esto, ya se conocen todos los elementos de la diagonal de la matriz $E(uu')$. Luego, la covarianza entre u_t y u_{t-1}, dado que la esperanza de u_t es cero, se calcula como:

$$cov(u_t u_{t-1}) = E(u_t u_{t-1})$$

Por lo tanto:

$$E(u_t u_{t-1}) = E[(\rho u_{t-1} + e_t)u_{t-1}] = E(\rho u_{t-1}^2 + e_t u_{t-1})$$

Introduciendo el operador esperanza:

$$E(u_t u_{t-1}) = \rho E(u_{t-1}^2) + E(e_t u_{t-1}) = \rho\sigma_u^2$$

Luego, la covarianza entre u_t y u_{t-2}:

$$E(u_t u_{t-2}) = E[(\rho u_{t-1} + e_t)u_{t-2}] = E(\rho u_{t-1}u_{t-2} + e_t u_{t-2})$$

Introduciendo el operador esperanza:

$$E(u_t u_{t-2}) = \rho E(u_{t-1}u_{t-2}) + E(e_t u_{t-2}) = \rho E(u_{t-1}u_{t-2})$$

Reemplazando:

$$E(u_t u_{t-2}) = \rho(\rho\sigma_u^2) = \rho^2\sigma_u^2$$

Finalmente para u_t y u_{t-3}:

$$E(u_t u_{t-3}) = E[(\rho u_{t-1} + e_t)u_{t-3}] = E(\rho u_{t-1}u_{t-3} + e_t u_{t-3})$$

Introduciendo el operador esperanza:

$$E(u_t u_{t-3}) = \rho E(u_{t-1} u_{t-3}) + E(e_t u_{t-3}) = \rho E(u_{t-1} u_{t-3})$$

Reemplazando:

$$E(u_t u_{t-3}) = \rho(\rho^2 \sigma_u^2) = \rho^3 \sigma_u^2$$

Recursivamente, se obtiene la expresión característica para cada covarianza, la cual viene dada por:

$$E(u_t u_{t-(T-1)} = \rho^{T-1} \sigma_u^2$$

Por lo tanto, la matriz de varianzas y covarianzas del término de error vendrá dada por:

$$E(uu') = \begin{pmatrix} \sigma_u^2 & \rho\sigma_u^2 & \cdots & \rho^{T-1}\sigma_u^2 \\ \rho\sigma_u^2 & \sigma_u^2 & \cdots & \rho^{T-2}\sigma_u^2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1}\sigma_u^2 & \rho^{T-2}\sigma_u^2 & \cdots & \sigma_u^2 \end{pmatrix}$$

Factorizando, se llega a lo siguiente:

$$E(uu') = \sigma_u^2 \begin{pmatrix} 1 & \rho & \cdots & \rho^{T-1} \\ \rho & 1 & \cdots & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \cdots & 1 \end{pmatrix}$$

Por lo tanto, en este contexto, conociendo ρ es posible conocer completamente la matriz de varianzas y covarianzas del término de error. De hecho, se ha asumido que es conocido, y por ende, se conoce la matriz Ω, con lo cual se puede aplicar la descomposición espectral, encontrar la matriz P, transformar el modelo y estimar por MCG. De hecho, en este caso particular para cuando el error es un proceso autorregresivo de primer orden, AR(1), la matriz P viene dada por:

$$P = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}$$

Dado que el modelo viene dado por:

$$y_t = x_t'\beta + u_t$$

y dado que la matriz P premultiplica el vector y, la primera observación del vector transformado (Py) quedaría como:

$$\sqrt{1-\rho^2} y_t = \sqrt{1-\rho^2} x_t'\beta + \sqrt{1-\rho^2} u_t$$

Y el resto de las observaciones ($n-1$), quedaría como:

$$y_t - \rho y_{t-1} = (x_t - \rho x_{t-1})'\beta + u_t - \rho u_{t-1}$$

Por lo tanto, el modelo transformado queda:

$$y_t^* = x_t^{*\prime}\beta + u_t^*$$

y el término de error está bien comportado (iid) pues:

$$u_t^* = u_t - \rho u_{t-1} = e_t$$

En caso de que la forma de la matriz Ω fuera conocida, como este caso, pero que el parámetro ρ fuera desconocido, entonces, habría que estimarlo, y la estimación del modelo se llevaría a cabo mediante el estimador de MCF. Asumiendo el mismo ejemplo anterior, el estimador de MCF se implementaría a través de los siguientes pasos: se estima el modelo original por MCO ($y_t = x_t'\beta + u_t$), con lo cual se obtienen los errores de MCO ($\hat{u}$). Luego, se corre una regresión auxiliar entre $\hat{u}_t$ y $\hat{u}_{t-1}$, con lo cual es obtiene una estimación para ρ ($\hat{\rho}$). Con ese valor para ρ se transforma el modelo original, y se corre la siguiente regresión:

$$y_t - \hat{\rho} y_{t-1} = (x_t - \hat{\rho} x_{t-1})'\beta + error_t$$

Así se obtiene una estimación para β, y con esa estimación para β ($\hat{\beta}$) se encuentran nuevamente los errores de MCO ($\hat{u}$), con lo cual se puede obtener una nueva estimación para ρ ($\hat{\rho}^{(2)}$). Se tranforma nuevamente el modelo de la siguiente forma:

$$y_t - \hat{\rho}^{(2)} y_{t-1} = (x_t - \hat{\rho}^{(2)} x_{t-1})'\beta + error_t$$

Así, se obtiene una nueva estimación para β ($\hat{\beta}^{(2)}$). Este procedimiento iterativo se repite hasta que las estimaciones de β converjan, es decir, hasta que:

$$\hat{\beta}^{(j)} \approx \hat{\beta}^{(j+1)}$$

Este procedimiento de estimación, que puede generalizarse para procesos AR de órdenes superiores se conoce como el método de Cochrane-Orcutt.

¿Qué hacer en caso de que la matriz Ω sea completamente desconocida? Newey-West (1987) desarrolló un mecanismo de estimación para cuando la matriz Ω es completamente desconocida. Para presentar dicho método, primero se debe recordar que en presencia de perturbaciones no esféricas, la matriz de varianzas y covarianzas del estimador MCO viene dada por:

$$var(\hat{\beta}) = \sigma^2 (X'X)^{-1} X'\Omega X (X'X)^{-1}$$

Luego, manipulando la expresión:

$$var(\hat{\beta}) = T(X'X)^{-1}\sigma^2 \frac{X'\Omega X}{T}(X'X)^{-1}$$

Newey-West (1987) proponen estimar la matriz:

$$S = \sigma^2 \frac{X'\Omega X}{T}$$

de la siguiente forma:

$$\hat{S} = \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T} w_{t-s}\hat{u}_t\hat{u}_s x_s x_s'$$

donde:

$$w = 1 - \left(\frac{t-s}{L+1}\right)$$

El parámetro L corresponde al máximo orden de autocorrelación del término de error, por lo tanto, la implementación de la corrección de Newey-West exige conocer el orden de autocorrelación del término de error.

Al igual que para el caso de la heterocedasticidad, existen varios test diseñados para detectar la presencia de la autocorrelación. A continuación se presentan los más utilizados.

Test de Durbin-Watson. Este test solo permite detectar la presencia de procesos AR(1) para el error, lo cual se convierte en su principal desventaja. Así, el test permite detectar si el error exhibe el siguiente comportamiento:

$$u_t = \rho u_{t-1} + e_t$$

La hipótesis nula es que los errores no están autocorrelacionados. El test se basa en el siguiente estadistico:

$$d = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T}\hat{u}_t^2}$$

Es directo mostrar que:

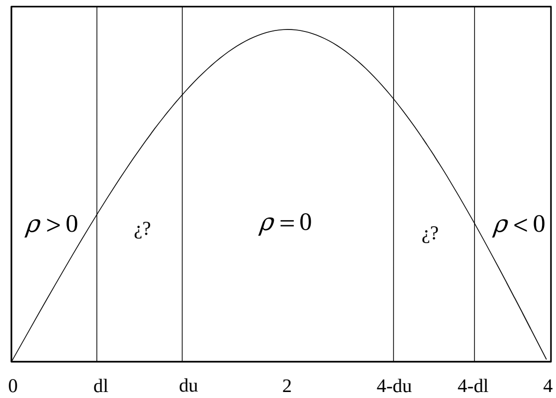
$$d \approx 2(1-\rho)$$

donde:

$$\rho = \frac{\sum_{t=2}^{2}\hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_{t-1}^2}$$

Luego, note que si $d < dl$, entonces, $\rho > 0$, es decir, habría evidencia para establecer la presencia de autocorrelación positiva de primer orden. Por otro lado, si $d = 2$, entonces, $\rho = 0$, es decir, no habría autocorrelación de primer orden. Y finalmente, si $d > 4 - dl$, entonces, $\rho < 0$, y habría evidencia a favor de la presencia de autocorrelación negativa de primr orden. La figura 19 resume estos resultados.

Figura 19: Test de Durbin-Watson



Para implementar el test primero se estima el modelo por MCO, y se obtienen los residuos de MCO. Con ellos se procede a estimar ρ y se calcula d. Dependiendo de la zona en la que caiga el valor de este estadístico se determina lo que indica la evidencia estadística. Las desventajas de este test son dos: primero, solo permite detectar procesos AR de primer orden, y segundo, el test tiene zonas de indeterminación en donde no es posible señalar qué es lo que está ocurriendo. En efecto, la distribución exacta del test (d) depende de las observaciones particulares de la matriz X. Sin embargo, los valores críticos de este test se encuentran dentro de dos distribuciones límites, lo cual da origen a zonas de indeterminación

Test h de Durbin. Este test es una mejora del anterior por cuando permite que en el modelo original esté la variable dependiente rezagada como un regresor adicional. En efecto, suponga que el modelo viene dado por:

$$y_t = x_t'\beta + \gamma y_{t-1} + e_t$$

Note que entre los regresores aparece la variable dependiente rezagada. Por cierto, pueden haber más rezagos de la variable dependiente. El test es el siguiente:

$$h = \left(1 - \frac{d}{2}\right)\sqrt{\left(\frac{T}{1 - Tvar(\hat{\gamma})}\right)}$$

donde $var(\hat{\gamma})$ es la varianza del coeficiente que acompaña al rezago de la variable dependiente en el modelo original. Este estadístico (h) se disribuye como una normal estándar, con lo cual se procede a evaluar la hipótesis nula (ausencia de autocorrelación).

Test de Breusch y Godfrey. La principal ventaja de este test, por sobre los anteriores, es que permite detectar órdenes superiores de autocorrelación. Los pasos mediante los cuales se implementa este test son los siguientes. En primer lugar se debe estimar el modelo original por MCO (pueden haber rezagos de la variable dependiente) con lo cual se obtienen los residuos de MCO $(\hat{u})$. A continuación se corre una regresión auxiliar entre $\hat{u}$ y, por ejemplo, p rezagos de los errores $(\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-p})$, más las otras variables explicativas del modelo original (x). Luego, se calcula el estadístico TR^2, donde el R^2 corresponde al coeficiente de determinación de la regresión auxiliar. Finalmente, si el valor del estadístico TR^2 (el cual se distribuye Chi-cuadrado con p grados de libertad) es mayor que el valor crítico de la distribución Chi-cuadrado, existe evidencia para rechazar la hipótesis nula de no autocorrelación.

Test de Box-Pierce-Ljung. Este test también sirve para detectar órdenes superiores de autocorrelación. El test se define de la siguiente manera:

$$Q = T \sum_{j=1}^{p} r_j^2$$

donde p es el orden de autocorrelación a testar, y:

$$r_j = \frac{\sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}}{\sum_{t=1}^{T} \hat{u}_t^2}$$

donde los errores $\hat{u}$ corresponden a los residuos de MCO del modelo original. El estadístico Q se distribuye como una Chi-cuadrado con p grados de libertad. Si el valor de Q excede el valor crítico de la distribución Chi-cuadrado, entonces se rechaza la hipótesis nula de no autocorrelación.

5. Problemas con los datos

Este último capítulo aborda la problemática que a veces surge al trabajar con datos reales. Se analizarán dos problemas que a menudo surgen en el análisis empírico, el de multicolinealidad, y el de error de medida.

5.1. Multicolinealidad

En términos concretos la multicolinealidad consiste en la correlación que existe entre las variables explicativas del modelo. Dado que las variables siempren tienen un cierto grado de correlación, la multicolinealidad es un problema de grado, y no de existencia. Por ejemplo, considere el siguiente modelo:

$$C_t = \alpha_1 + \alpha_2 I_t + \alpha_3 r_t + e_t$$

Como se puede apreciar, esta ecuación representa la función consumo keynesiana en donde el consumo (C) depende del ingreso (I), de la tasa de interés (r), y de otros factores no observables que se resumen en un shock estocástico (e). Es directo notar que el ingreso y la tasa de interés están correlacionados negativamente, y por ende, que existe un problema de multicolinalidad en este modelo. Ahora bien, ¿qué consecuencias tiene este hecho? Para ver los efectos que produce la presencia de multicolinealidad, considere el modelo de regresión lineal en su expresión matricial, $Y = X\beta + u$. En este caso, el estimador MCO viene dado por $\hat{\beta} = (X'X)^{-1}X'Y$, y su varianza es, $var(\hat{\beta}) = \sigma^2(X'X)^{-1}$. Por lo tanto, en la medida que exista un grado de correlación en las variables explicativas del modelo, la matrix $X'X$ tenderá a la singularidad, es decir, su determinante se hará más pequeño, la inversa sera cada vez más grande, $(X'X)^{-1}$, afectando así las estimaciones de MCO. Así, al ser la matriz de varianzas y covarianzas más grande, se producirá una pérdida de eficiencia en las estimaciones de MCO. En la medida que el grado de correlación sea razonable, será posible seguir estimando el modelo por MCO, sin embargo, si la correlación es perfecta (regresores colineales), entonces, no será posible calcular la inversa de la matriz $X'X$, pues el determinante de ésta ($|X'X|$) tenderá a cero. En efecto, recuerde que:

$$(X'X)^{-1} = \frac{1}{|X'X|} adj(X'X)$$

Por lo tanto, en la medida que la multicolinealidad no sea exacta, es decir, que los regresores tengan cierto grado de correlación, pero que no sean colineales, el modelo se seguirá estimando por MCO. No obstante, mientras mayor sea el grado de correlación, más ineficiente serán las estimaciones, lo que en sí puede constituir un problema toda vez que los test t serán cada vez más pequeños, y por ende, los coeficientes estimados puedieran no ser estadísticamente significativos; no porque no lo sean desde un punto de vista económico, sino más bien por un problema estadístico. Cuando la multicolinealidad es exacta, es decir, algunos regresores son colineales, entonces no se podrá estimar por MCO ya que la matriz $X'X$ no será invertible. Este es precisamente el caso cuando por ejemplo en un modelo se incorporan variables dummies para cada una de las categorías de género, además de una constante. Tal como se vió anteriormente, esto hace que la primera columna de la matriz X sea una combinación lineal exacta de las variables dummies para hombre y mujer.

Para ver más claramente el efecto que tiene la multicolinealidad en la varianza de los coeficientes estimados, considere la siguiente partición de la matriz X:

$$X = \begin{pmatrix} x_j & X_j \end{pmatrix}$$

donde x_j es el vector columna del j-ésimo regresor. Por lo tanto:

$$X'X = \begin{pmatrix} x_j'x_j & x_j'X_j \\ X_j'x_j & X_j'X_j \end{pmatrix}$$

Luego, la inversa de esta matriz viene dada por:

$$(X'X)^{-1} = \begin{pmatrix} (x_j'x_j - x_j'X_j(X_j'X_j)^{-1}X_j'x_j)^{-1} & ? \\ ? & ? \end{pmatrix}$$

Note que solo se ha presentado el elemento (1,1) de la inversa de la matriz $(X'X)$ pues se quiere encontrar una expresión para la varianza del j-ésimo regresor del modelo. Manipulando dicho término se tiene que:

$$(x_j'x_j - x_j'X_j(X_j'X_j)^{-1}X_j'x_j)^{-1} = (x_j'(I - X_j(X_j'X_j)^{-1}X_j')x_j)^{-1}$$

Por lo tanto:

$$(x_j'x_j - x_j'X_j(X_j'X_j)^{-1}X_j'x_j)^{-1} = (x_j'M_jx_j)^{-1}$$

Es decir, la varianza para el coeficiente del j-ésimo regresor viene dada por:

$$var(\hat{\beta}_j) = \sigma^2(x_j'M_jx_j)^{-1} = \frac{\sigma^2}{x_j'M_jx_j}58 \tag{60}$$

Es posible mostrar que:

$$var(\hat{\beta}_j) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_{ji} - \bar{x_j})^2(1 - R_j^2)}$$

donde R_j^2 es el coeficiente de determinación de una regresión entre el j-ésimo regresor contra el resto de las variables explicativas (X_j). Por lo tanto, es posible apreciar que la varianza del estimador de MCO para β_j depende de tres factores: (i) de la varianza del término de error (σ^2), (ii) de la variación total de la variable en cuestión, y (iii) de la relación que existe entre la variable x_j y el resto de las variables explicativas X_j representada por el R_j^2. Note que mientras mayor sea la multicolinealidad entre x_j y el resto de las variables explicativas (es decir, mientras mayor sea el R_j^2), mayor será también la varianza estimada para $\hat{\beta}_j$. Es decir, la multicolinealidad genera una pérdida de eficiencia. Más formalmente:

$$\frac{\partial var(\hat{\beta}_j)}{\partial R_j^2} > 0$$

Dado que la multicolinealidad no es un problema grave en la medida que no sea exacta, hay econometristas que no le asignan mucha importancia a esta problemática. No obstante, hay quienes sí le dedican algo más de cuidado, y es por esto que se han desarrollado algunos métodos para detectar la presencia de multicolinealidad. Un indicio de presencia de multicolinealidad consiste es cuando el modelo estimado presenta un elevado R^2, y los coeficientes estimados $(\hat{\beta})$ no son estadísticamente significativos. Esto se debe al hecho de que al estar tan correlaciones los regresores, no es posible identificar el efecto individual que cada uno tiene sobre la variable dependiente, pero en conjunto sí son capaces de explicar la varianza de la variable dependiente. Es por esta razón que se tiene un modelo con un elevado coeficiente de determinación pero con coeficientes que no son estadísticamente signficiativos a nivel individual. Otro indicio de la presencia de multicolinealidad en los datos es que pequeños cambios en los datos (por ejemplo, al agregar una nueva observación a los datos) se producen grandes cambios en los coeficientes estimados $(\hat{\beta})$. Asimismo, cuando los coeficientes estimados tienen signos distintos a los esperados, o una magnitud poco creíble (recuerde que la multicolinealidad afecta tanto a los coeficienes de MCO como a sus varianzas) también es un indicio de la existencia de multicolinealidad.

Los métodos anteriores no son muy formales, sin embargo, Belsley propuso un test que sirve para determinar más objetivamente cuándo la multicolinealidad es un problema grave, en otras palabras, sirve para determinar el grado en que está presente la multicolinealidad en el modelo. Belsley propone construir la siguiente matriz:

$$B = S(X'X)S$$

La matriz S se construye de la siguiente manera:

$$S = \begin{pmatrix} \frac{1}{\sqrt{x_1'x_1}} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sqrt{x_2'x_2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sqrt{x_k'x_k}} \end{pmatrix}$$

donde el término x_h corresponde a la h-ésima columna de la matriz X. Luego, Belsley propone calcular:

$$\gamma = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$$

donde λ_{max} y λ_{min} son los valores propios, máximo y mínimo respectivamente, de la matriz S. el test establece que si $\gamma > 20$ entonces, existe un problema grave de multicolinealidad. La siguiente cuestión consiste en determinar qué hacer cuándo se detecta un problema grave en términos de multicolinealidad. Las opiniones en este aspecto están dvididas. Algunos establecen que, dado que no se pueden modificar las características de los datos, entonces lo mejor es no hacer nada en la medida que la multicolinealidad no sea exacta. Por otro lado, si existe un grado de multicolinealidad, y se opta por eliminar el regresor que la provoca, se puede generar un problema aún más grave que es la omisión de variable relevante. Ahora bien, si la multicolinealidad es exacta, entoces, no se puede estimar por MCO, y por ende, habría que eliminar el regresor que está provocando el problema. Otra alternativa consiste en implementar el estimador de Ridge. El estimador de Ridge se define de la siguiente manera:

$$\hat{\beta}_R = (X'X + rD)^{-1}X'y$$

donde $D = diag(X'X)$, y r es un escalar cualquiera. Al introducir un valor a cada elemento de la matriz $X'X$, ahora ésta es invertible, y por ende, se pueden encontrar los estimdores de β. Sin embargo, se debe tener presente que el estimador de Ridge es sesgado, pero tiene una maner varianza.

Dado que los remedios para la multicolinealidad afectan los parámetros que originalmente el econometrista desea estimar entonces hay quienes optan por no hacer nada. En la medida que la multicolinealidad no sea exacta siempre será posible estimar por MCO, aunque las varianzas serán menos eficientes.

5.2. Error de medida

Otro problema que es frecuente en el trabajo empírico guarda relación con el hecho de que a veces las variables no están correctamente medidas y tienen errores de medición. Por ejemplo, suponga que se desea estimar la relación existente entre los salarios de las personas y los años de escolaridad. Los salarios podrían tener error de medida pues a veces las personas señalan un salario aproximado y no indican el salario exacto que perciben en sus empleos. Por ello cuando el investigador desee estimar el modelo estará utilizando información que tiene error de medida. Ahora bien tanto la variable dependiente como la variables independientes podrían tener error de medida, y por ello a continuación se analizan las consecuencias que esto tendrá en las estimaciones de MCO.

En primer lugar, suponga que es la variable dependiente la que está medida con error. Por ejemplo, suponga que la variable y (que se encuentra medida en desvios con respecto a su media) viene dada por:

$$y = y^* + e$$

En este contexto, y^* es el verdadero valor de la variable dependiente, sin embargo, el econometrista observa y, la cual está medida con error. Además, se asume que e es un shock estocástico independiente e identicamente distribuido con media cero y varianza σ_e^2. Considere ahora el modelo de regresión lineal simple, el cual también se encuentra en desvíos con respecto a la media (por eso no contiene constante):

$$y_i^* = \beta x_i + u_i$$

donde u se distribuye iid con media cero y varianza σ_u^2. Sin embargo, el econometrista no observa y^* sino que observa y. Dado que $y = y^* + e$, reemplazando en la ecuación anterior se tiene que:

$$y_i - e_i = \beta x_i + u_i$$

Luego:

$$y_i = \beta x_i + u_i + e_i = \beta x_i + w_i$$

donde $w = u + e$ es el nuevo término de error, que es iid con media cero y $var(w) = \sigma_w^2 = \sigma_u^2 + \sigma_e^2$.[76] Este es el modelo que estimará el econometrista, pues está constituido por las variables (x e y) que efectivamente observa. El estimador de MCO de β se calcula como:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Reemplazando y:

$$\hat{\beta} = \frac{\sum x_i(\beta x_i + u_i)}{\sum x_i^2} = \beta + \frac{\sum x_i u_i}{\sum x_i^2}$$

Aplicando el operador esperanza se llega a:

$$E(\hat{\beta}) = \beta + \frac{\sum x_i E(u_i)}{\sum x_i^2} = \beta$$

Por lo tanto, cuando la variable dependiente está medida con error el estimador de MCO será insesgado y consistente. Sin embargo, dado que $\sigma_w^2 > \sigma_u^2$, la varianza de $\hat{\beta}$ será mayor que en ausencia de error de medida.

Las cosas son distintas cuando es la variable independiente la que está medida con error. En efecto, ahora asuma que la variable dependiente no está medida con error, pero que la variable independiente si lo está. En efecto, suponga que:

$$x_i = x_i^* + v_i$$

[76] El shock w hereda las propiedades estadísticas de u y e.

donde la variables x y x^* están medidas en desvíos con respecto a su media. Es decir, el econometrista observa x_i, la cual está medida con error, en lugar de x_i^*. El shock v_i es iid con media cero y varianza σ_v^2. El modelo nuevamente viene dado por (en desvíos con respecto a la media):

$$y_i = \beta x_i^* + u_i$$

Reemplazando x^*:

$$y_i = \beta(x_i - v_i) + u_i = \beta x_i + (u_i - \beta v_i) = \beta x_i + \zeta_i$$

Recuerde que para que el estimador de MCO sea insesgado, se debe cumplir que la covarianza entre la variable independiente (x) y el error del modelo (ζ) sea igual a cero. Como se verá a continuación este no es el caso. En efecto, note que:

$$cov(x,\zeta) = cov(x^* + v, u - \beta v) = cov(x^*, u) + cov(x^*, -\beta v) + cov(v, u) + cov(v, -\beta v)$$

Note que todas las covarianzas que contienen un shock iid son cero, por lo tanto:

$$cov(x,\zeta) = cov(v, -\beta v) = -\beta cov(v, v) = -\beta var(v) = -\beta\sigma_v^2$$

Dado que la $cov(x,\zeta) \neq 0$, el estimador MCO de β es sesgado. El estimador MCO viene dado por:

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$$

Reemplazando y:

$$\hat{\beta} = \frac{\sum x_i(\beta x_i^* + u_i)}{\sum x_i^2} = \beta\frac{\sum x_i x_i^*}{\sum x_i^2} + \frac{\sum x_i u_i}{\sum x_i^2}$$

Al aplicar el operador esperanza es directo comprobar que $E(\hat{\beta}) \neq \beta$. Para analizar si el estimador de MCO en este contexto es consistente, se reemplaza x en el numerador del primer término:

$$\hat{\beta} = \beta\frac{\sum(x_i^* + v_i)x_i^*}{\sum x_i^2} + \frac{\sum x_i u_i}{\sum x_i^2}$$

Desarrollando la expresión anterior:

$$\hat{\beta} = \beta\frac{\sum x_i^{*2}}{\sum x_i^2} + \beta\frac{\sum x_i^* v_i}{\sum x_i^2} + \frac{\sum x_i u_i}{\sum x_i^2}$$

Multiplicando numeradores y denominadores por $1/n$:

$$\hat{\beta} = \beta\frac{(1/n)\sum x_i^{*2}}{(1/n)\sum x_i^2} + \beta\frac{(1/n)\sum x_i^* v_i}{(1/n)\sum x_i^2} + \frac{(1/n)\sum x_i u_i}{(1/n)\sum x_i^2}$$

Aplicando *plim*:

$$plim(\hat{\beta}) = \beta\frac{plim(1/n)\sum x_i^{*2}}{plim(1/n)\sum x_i^2} + \beta\frac{plim(1/n)\sum x_i^* v_i}{plim(1/n)\sum x_i^2} + \frac{plim(1/n)\sum x_i u_i}{plim(1/n)\sum x_i^2}$$

Dado que el $plim(1/n)\sum x_i^* v_i$ y el $plim(1/n)\sum x_i u_i$ son cero pues corresponden a las covarianzas poblacionales de x^* y v, y de x y u, la expresión anterior se reduce a:

$$plim(\hat{\beta}) = \beta \frac{plim(1/n)\sum x_i^{*2}}{plim(1/n)\sum x_i^2} = \beta\left(\frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_v^2}\right) < 1$$

donde $\sigma_{x^*}^2$ es la varianza de X^*. Por lo tanto, $plim(\hat{\beta}) < \beta$, es decir, el estimador MCO de β subestima al verdadero parámetro poblacional. A este sesgo asintótico se le conoce como sesgo de atenuación.

En resumen, cuando la variable dependiente está medida con error el estimador de MCO sigue siendo insesgado y consistente, sin embargo, cuando las variables independientes están medidas con error el estimador de MCO será inconsistente. ¿Qué hacer en este caso?

El método de estimación de variables instrumentales (VI) ofrece una solución.[77] Primero se debe definir en qué consiste un instrumento. Un instrumento (Z) es una variable que cumple con las siguientes dos propiedades:

1. Exogeneidad: $cov(Z, u) = 0$
2. Relevancia: $cov(Z, X) \neq 0$

Es decir, es una variable que no está correlacionada con el término de error del modelo, y está fuertemente correlacionada con la variable del modelo que presenta el problema, en este caso la variable X que está medida con error. Luego, el estimador de VI se implementa en dos etapas. En la primera etapa se estima una regresión entre la variable con problemas (X), y el instrumento (Z). De es regresión es posible calcular la proyección de X ($\hat{X}$). Luego, en una segunda etapa se corre una regresión entre la variable dependiente del modelo original (Y) contra $\hat{X}$, de donde se obtiene el estimador de VI ($\hat{\beta}_{VI}$). Si se tienen varios instrumentos, entonces, el método se aplica corriendo en la primera etapa la regresión entre la variable con problemas (X) contra todos los instrumentos disponibles ($Z_1, Z_2, \ldots, Z_q$), con lo cual se obtiene $\hat{X}$. Cabe señalar que el estimador de VI es consistente, pero no es insesgado. Además, es menos eficiente que el estimador de MCO, es decir:

$$var(\hat{\beta}_{VI}) > var(\hat{\beta}_{MCO})$$

Tal como se analizó, cuando las variables independientes del modelo están medidas con error, el estimador de MCO es inconsistente, mientras que el estimador de variables instrumentales es consistente. Ahora bien, si en realidad no hubiese errores de medida, entonces ambos estimadores serían consistentes, y MCO además sería eficiente, lo cual no ocurre con el estimador de variables instrumentales pues es un estimador que se implementa dos etapas, lo que le hace perder eficiencia. Por eso,

[77] Este método sirve para cualquier caso en que $cov(X, u) \neq 0$. Por ejemplo, cuando se analizó el tópico de omisión de variable relevante se vio el caso de la estimación de una ecuación de salario. En ese caso cuando el econometrista omitía la habilidad del individuo se producía sesgo e inconsistencia pues $cov(S, u) \neq 0$. El método de VI permite enfrentar también ese caso.

para testar si las variables independientes están medidas con error y están produciendo inconsistencia, se puede implementar el test de Hausman (1978), el cual viene dado por:

$$H = (\hat{\beta}_{MCO} - \hat{\beta}_{VI})'(var(\hat{\beta}_{VI}) - var(\hat{\beta}_{MCO}))^{-1}(\hat{\beta}_{MCO} - \hat{\beta}_{VI})$$

el cual se distribuye Chi-cuadrado con g grados de libertad, siendo g el número de regresores que potencialmente estarían medidos con error. La hipótesis nula del test es que no existe error de medida, es decir, $H_0 : \hat{\beta}_{MCO} = \hat{\beta}_{VI}$. Luego, cuando el valor para H excede el valor crítico de la distribución Chi-cuadrado, se rechaza la hipótesis nula, y el estimador MCO es inconsistente.

Ejercicios propuestos

A. Comentes

Comente cada una de las siguientes afirmaciones indicando claramente si éstas son verdaderas, falsas o inciertas. Justifique apropiadamente cada una de sus respuestas.

1. En el modelo de regresión lineal simple ($Y_i = \beta_1 + \beta_2 X_i + u_i$) no es deseable que la varianza de la variable explicativa sea alta, ya que provoca una mayor imprecisión en la estimación de mínimos cuadrador ordinarios (MCO).

2. El test t permite determinar si una variable explicativa tiene o no un efecto causal sobre la variable dependiente.

3. El R^2 es un estadístico que busca mediar la precisión de los estimadores de MCO.

4. Para que el estimador de MCO sea insesgado se requiere que el término de error se distribuya de manera normal.

5. El estimador de MCO presenta el menor error cuadrático medio (ECM) de entre todos los estimadores lineales e insesgados.

6. El estimador de MCO minimiza el R^2 de la regresión.

7. Siempre es posible obtener los estimadores de MCO en el modelo múltiple utilizando la expresión $\hat{\beta} = (X'X)^{-1}X'Y$.

8. La presencia de multicolinealidad hará que los test de significancia individual sean menores a los verdaderos debido al sesgo a la baja en la estimación de los parámetros.

9. La presencia de multicolinealidad no afectará las ecuaciones normales de las que se obtiene el estimador de mínimos cuadrados ordinarios.

10. El supuesto de normalidad de los residuos es clave en la obtención del test de Wald.

11. Se ha estimado un modelo de regresión simple que busca explicar la tasa de inversión en capital humano (Y) como función del gasto de gobierno (X) con una muestra que contiene datos de 100 países para el año 2008. Para confirmar qué tan robusto es el modelo, es necesario realizar un test de autocorrelación de los residuos.

12. Un elevado grado de correlación entre las variables explicativas del modelo provoca inconsistencia en los estimadores de mínimos cuadrados ordinarios.

13. La corrección de Newey-West hace que el estimador de MCO vuelva a ser MELI (mejor estimador lineal e insesgado).

14. No existen costos de agregar variables adicionales a un determinado modelo. Los estimadores MCO de los parámetros que sí pertenecen al verdadero modelo seguirán siendo consistentes e insesgados en presencia de variables irrelevantes.

15. Elegir un modelo que maximice el R^2 ajustado es equivalente a elegir un modelo que minimice el $\widehat{\sigma}^2$.

16. Un analista luego de estimar su modelo por MCO, obtiene un p-value para el test Jarque-Bera igual a 0.001. El investigador afirma entonces que no es posible trabajar con MCO ya que entrega resultados sesgados.

17. En un modelo que presenta una alta multicolinealidad, y en donde, a causa de ello, los estimadores no son estadísticamente significativos, se afirma que el estimador MCO ya no es MELI.

18. Si el objetivo es realizar una buena predicción, lo mejor es escoger aquel modelo que tenga un mayor R^2 asociado.

19. El estimador de máximo verosimil será eficiente solo en caso de que los residuos sean esféricos.

20. La propiedad de invarianza de los estimadores de máximo verosimiles nos asegura que si el estimador de θ es $\widehat{\theta}$ entonces el estimador de logaritmo de θ será la exponencial de θ.

21. Sólo si el término de error se distribuye normal es posible realizar inferencia estadística sobre los parámetros estimados.

22. Para realizar predicciones siempre es mejor utilizar un solo modelo en lugar de varios.

23. Si usted conoce la información poblacional no tiene sentido trabajar con intervalos de confianza.

24. Los residuos (errores) del modelo de regresión lineal se asumen normales pero no independientes e idénticamente distribuidos.

25. Para que el estimador de MCO sea insesgadp, es necesario que el error se distribuya normal con media cero y varianza constante para todas las observaciones.

26. Con respecto a la varianza que pueda tener la estimación de un modelo lineal simple, es preferible tener varianzas pequeñas tanto de la variable dependiente como de los regresores.

27. A mayor R^2 mejor es la calidad del modelo estimado.

28. Estimar β en $Y_i = \alpha + \beta X_i + u_i$ ó γ en $Y_i - \overline{Y} = \gamma(X_i - \overline{X}) + u_i$ genera el mismo resultado.

29. La omisión de una variable irrelevante hace que el estimador de MCO sea sesgado pero consistente.

30. Cuando el error se distribuye normal con media cero y varianza homocedástica el estimador de MCO es equivalente al estimador de máxima verosimilitud.

31. Si el test de White arroja evidencia de errores heterocedásticos, entonces, lo mas recomendable es implementar el estimador de mínimos cuadrados factibles (MCF), ya que al aplicar el estimador de MCO se obtendraán parámetros sesgados.

32. Cuando los regresores son estocásticos basta con que $E(u) = 0$ para que el estimado MCO sea insesgado.

33. Siempre que se omite una variable relevante el estimador de MCO será sesgado e inconsistente.

34. En presencia de multicolinealidad el estimador de MCO será sesgado pero consistente.

35. Cuando la variable dependiente está medida con error, el estimador de MCO será menos eficiente.

B. Ejercicios prácticos

1. Considere los siguientes datos:

X	2	3	1	5	9
Y	4	7	3	9	17

El modelo que relaciona las variables X e Y es el siguiente:

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

donde u es un término de error bien comportado.

a) Encuentre los estimadores de mínimos cuadrados ordinarios (MCO) para β_0 y β_1

b) Encuentre la matriz de varianzas y covarianzas de los estimadores MCO. Utilice el estimador insesgado para σ^2.

c) ¿Son estadísticamente significativos los coeficientes estimados?

d) Calcule R^2 y el R^2 ajustado.

Ahora alguien le informa que la matriz de varianzas y covarianzas del término de error viene dada por:

$$E(uu') = \sigma^2 \begin{pmatrix} 0.10 & 0 & 0 & 0 & 0 \\ 0 & 0.05 & 0 & 0 & 0 \\ 0 & 0 & 0.20 & 0 & 0 \\ 0 & 0 & 0 & 0.30 & 0 \\ 0 & 0 & 0 & 0 & 0.15 \end{pmatrix}$$

e) Calcule el estimador de mínimos cuadrados generalizados (MCG).

f) En base a este estimador, calcule el estimador insesgado de σ^2.

g) Calcule nuevamente la matriz de varianzas y covarianzas del estimador de MCO (Ayuda: no olvide tomar en cuenta la información relativa a la matriz de varianzas y covarianzas del término de error).

h) ¿Qué estimador es más eficiente, MCO o MCG? ¿Cómo lo demostraría? Explique.

2. Considere el siguiente proceso autoregresivo para los residuos:

$$\epsilon_t = \rho\epsilon_{t-2} + \upsilon_t$$
$$\epsilon = \upsilon_0$$
$$t = 1, \ldots, T$$
$$\upsilon_t \sim iidN(0, \sigma^2)$$
$$|\rho| < 1$$

a) Muestre que si t es par, entonces ϵ_t se puede escribir como una combinación lineal de los υ_t pares. Muestre que lo mismo se da para los t impares.

b) ¿Cuál es la varianza del proceso ϵ_t?

c) Encuentre las autocovarianzas de orden 1 y 2 de ϵ_t.

3. Considere la siguiente función de densidad de probabilidad para la variable aleatoria y:

$$f(y) = \frac{\left(\frac{\alpha}{\lambda}\right)^{\alpha-1} e^{-y/\lambda}}{\lambda(\alpha-1)}$$

con $0 < y < \infty$, $\alpha > 0$, y $\alpha \in N$. Se sabe que $E(y) = \alpha\lambda$.

a) Encuentre la función de log-likelihood.

b) Encuentre el estimador de máxima verosimilitud (MV) para λ.

c) Encuentre la varianza de $\hat{\lambda}_{MV}$. ¿Qué propiedad tiene esta varianza? Explique.

d) Considere ahora la siguiente información para la variable $y : y = \{2, 8, 4, 7, 9\}$ Suponiendo que $\alpha = 3$, encuentre el estimador de MV para λ.

e) Mediante el test de razón de verosimilitud (LRT) testee la siguiente hipótesis nula:

$$H_0 : \lambda = 3,45$$

4. Considere el siguiente modelo:

$$y_i = \alpha + u_i$$

con i=1,2,...,n.

a) Bajo el supuesto de que los errores son independientes e idénticamente distribuidos con media cero y varianza σ^2, demuestre que el estimador de mínimos cuadrados ordinarios $\widehat{\alpha}_{MCO}$ es igual al promedio de y. Calcule su varianza.

b) Asuma ahora que $E(uu') = \sigma\Omega$, donde Ω es una matriz diagonal, cuyo elemento i-ésimo corresponde a $w_i = x_i^2$. Demuestre que el estimador MCG de α es igual a:

$$\widehat{\alpha}_{MCG} = \frac{\sum_{i=1}^{n} \left(\frac{y_i}{x_i^2}\right)}{\sum_{i=1}^{n} \left(\frac{1}{x_i^2}\right)}$$

c) Calcule la varianza de $\hat{\alpha}_{MCG}$.

d) Bajo el mismo esquema de la pregunta b, cacule el estimador de MCO de α y su varianza.

e) Bajo el supuesto esgrimido en b, ¿qué estimador posee la menor varianza, MCO o MCG?

5. Considere el siguiente modelo:

$$Y_t = \alpha_1 + \alpha_2 t + e_t$$

donde t denota el tiempo e_t es una perturbación esférica con media cero (que no se distribuye necesariamente normal). Suponga que usted cuenta con los siguientes datos:

t	1	2	3	4	5	6
Y	-2	-2	0	1	2	3

a) Calcule el estimador de mínimos cuadrador ordinarios (MCO) de α_2 ($\hat{\alpha}_2$).

b) ¿Es insesgado este estimador? Demuestre.

c) Utilizando el estimador insesgado de la varianza del error calcule la varianza de $\hat{\alpha}_2$.

Considere ahora el siguiente estimador alternativo para α_2:

$$\tilde{\alpha}_2 = \frac{1}{8}(Y_6 + Y_5 - Y_2 - Y_1)$$

d) ¿Es insesgado este estimador alternativo? Demuestre.

e) Derive una expresión para la varianza de $\widetilde{\alpha}_2$.

f) Encuentre el error cuadrático medio (EMC) de $\widetilde{\alpha}_2$. Ayuda: El ECM viene dado por:

$$ECM(\widetilde{\alpha}_2) = E(\widetilde{\alpha}_2 - \alpha_2)$$

g) ¿Qué estimador tiene la menor varianza, $\hat{\alpha}_2$ o $\widetilde{\alpha}_2$?

6. En la producción de bisbirulitos cuadrados, variedad tropical de los bisbirulitos redonde, se utiliza como único insumo el trabajo. En esta indsutria, no existen ningún tipo de barreras a la entrada o a la salid y el trabajo utilizado no es específico. Existen razones teóricas para supone que la función de producción es de la forma $Y = \alpha L^b$, donde L es una medida de la cantidad de trabajo utilizado por unidad de tiempo, a y b son constentes e Y es la producción de bisbirulitos por unidad de tiempo. Usted desea testar la hipótesis nula de competencia perfecta,según la cual la función de producción debería estar homogénea de grado uno. Para ello, usted dispone de una base de datos donde se encuentra, para el año 1998, la producción y la cantidad utilizada de trabajo de cada una de las empresas que constituyen la industria. Para poder realizar el test de hipótesis, usted debe en primer término estimar a y b. Para ello, usted enfrenta dos problemas: primero, tal como está planteado, el modelo es no lineal; segundo, debe pasar del modelo matemático de la teoría al modelo econométrico, lo que implica introducir aleatoriedad en la especificación.

   a) Dado que usted sólo sabe estimación lineal, ¿existe alguna transformación que pueda hacerse de forma tal que pueda estimarse el modelo por MCO?

   b) ¿Cómo pasaría usted del modelo matemático al modelo econométrico? o lo que es lo mismo, ¿cómo introduciría un componente aleatorio?

   c) De existir la transformación considerada en (a), ¿cuál es la interpretación de los parámetros de a y b en el modelo transformado?

d) Establezca la hipítesis nula de la competencia perfecta. ¿Cómo testaría esta hipótesis? Responda especificando el estadígrafo a utilizar, su distribución bajo la nula, el cireterio de rechazo o no rechazo y los grados de libertad.

7. Considere el siguiente modelo

$$y_t = \beta_1 + \beta_2 x_t + \epsilon_t$$
$$\epsilon_t = \rho\epsilon_{t-1} + \mu_t$$
$$\mu_t \sim iidN(0, \sigma_\mu^2)$$

a) Demuestre que la varianza del término de error viene dada por:

$$var(\epsilon_t|x_t) = \sigma_t^2 = \frac{\sigma_\mu^2}{1-\rho^2}$$

b) Demuestre que:

$$cov(\epsilon_t, \epsilon_{t-j}|x_t, xt-j) = \rho^j \left(\frac{\sigma_\mu^2}{1-\rho^2}\right)$$

c) ¿Qué importancia tiene el supuesto de estacionariedad para el cálculo de las varianzas y covarianzas? Explique.

8. Considere el siguiente modelo:

$$Y_i = \alpha + \epsilon_i$$

donde:

$$E(\epsilon_i) = 0$$
$$var(\epsilon_i) = \sigma^2 x_i^2$$

a) Encuentre la matriz Ω.

b) Encuentre el estimador MELI de α. ¿Qué puede comentar de este resultado?

c) Encuentre el estimador de la varianza de α.

9. En Singapur los analistas financieros de distintas corredoras de bolsa y fondos de inversión (como en otras parte del mundo) han observado que el retorno de las acciones tiende a tener retornos llamativamente altos los días viernes y considerablemente más bajos los lunes. Asimismo, plantean que los retornos de los meses de Enero son siempre más altos que los otros meses. El modelo con el que ellos intentan predecir el retorno de las acciones se conoce en la literatura

financiera como el modelo CAPM, el cual se expresa epor una sencilla ecuación teórica:

$$RA_t = RF_t + \beta(RM_t - RF_t) \Rightarrow RA_t - RF_t = \beta(RM_t - RF_t)$$

Esta ecuación indica que RA es el retorno de la acción A en el momento t y se explica por la tasa libre de riesgo RF en el momento t y el 'premio por riesgo' o la brecha entre el retorno del portafolio de mercado y la tasa libre de riesgo $(RM - RF)$ ponderado por β que es el factor que indica el riesgo marginal de incluir la acción A en el portafolio. Un grupo de analistas de la empresa *Singapur Investments Corporation* ha trabajado actualizando un modelo tipo CAPM por varios años. En este modelo incorporan aleatoriedad y un término intercepto (C) para poder obtener un R^2 no negativo con certeza.

La última estimación arrojó con datos diarios de 1990 y 2000 los siguientes resultados:

Dependent Variable:RA-RF
Method Least Square
Sample 01-01-1999 12-29-2000
Included observations: 521

Variable	**Coefficient**	**Std.Error**	**t-Statistic**	**Prob.**
C	0.00037	0.003312	0.111819	0.911
RM-RF	0.796852	0.085907	9.275804	0
R-squared	0.142206	Mean dependent var		0.029211
Adjusted R-squared	0.140553	S.D dependent var		0.02811
S.E of regression	0.02606	F-statistic		86.04054
Sum squared resid	0.35247	Prob(F-statistic)		0.000000

a) Un financista que vio el trabajo de su colega comentó: "Con este modelo de R^2 tan bajo no podemos predecir nada. Piensalo: Sólo explicamos el 14 por ciento de la variación del retorno de la acción. Un modelo de R^2 tan bajo no tiene casi poder explicativo". Comente y fundamente a favor o en contra de esta afirmación con sustento estadístico.

b) Otro colega vio la posibilidad de incoporar los efectos "Día Lunes", "Día Viernes", "Enero" a través de la introducción de variables dummies (*Dlun*, *Dvie* y *Dene*). Con ellas estimó el siguiente modelo:

Dependent Variable:RA-RF
Method Least Square
Sample 01-01-1999 12-29-2000
Included observations: 521

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	-0.002171	0.003396	-0.639345	0.5229
RM-RF	0.822357	0.084015	9.788202	0.0000
DLUN	-0.006571	0.002874	-2.2866056	0.0227
DVIE	0.008948	0.002871	3.116289	0.0019
DENE	0.013979	0.004085	3.421915	0.0007
R-squared	0.191556	Mean dependent var		0.029211
Adjusted R-squared	0.185289	S.D dependent var		0.028110
S.E of regression	0.025373	Akaike info creiterion		-4.500723
Sum squared resid	0.332192	Schwarz criterion		-4.459881
Log likelihood	1177.438	F-statistic		30.565850
Durbin-Watson stat	1.863535	Prob(F-statistic)		0.000000

Con estos antecedentes determine con qué modelo se quedaría, justifique su respuesta.

c) ¿Cuánto retorno esperaría adicionalmente un lunes de Enero?

d) Explique cómo testaría las siguientes hipótesis: (Sólo se pide la hipótesis nula y el test estadístico que utilizaría, no realice cálculos).

- El aumento del retorno de la acción A un día viernes se compensa con la caída del retorno en un día lunes.
- El retorno de una acción un lunes de enero equivale al incremento del retorno un día viernes cualquiera de otro mes del año.

10. Considere la siguiente variable binominal, que toma valores de 0 y 1 de acuerdo a la siguiente función de distribición de probabilidad:

$$f(y) = \theta^y (1-\theta)^{(1-y)}$$
$$0 \leq \theta \leq 1$$
$$y = 0, 1$$

Así, la probabilidad de éxito ($y = 1$) es $f(1) = \theta$, y la probabilidad de fracaso ($y = 0$) está dada por $f(0) = 1 - \theta$.

a) Verificar que $E(y) = \theta$ y $var(y) = 1 - \theta$.

b) Asumiendo que se dispone de una muestra aleatoria de n observaciones provenientes de esta distribución, encuentre el estimado de MV de θ.

c) Encuentre la varianza asintótica del estimado de MV.

11. Considere el siguiente modelo:

$$Y = X\beta + u$$

donde Y es un vector de dimensión nx1, X es una matriz de nxk, y β es un vector de parámetros (desconocidos) de dimensión $kx1$. Asuma que el término de error (u) está bien comportado, y se distribuye independiente e idénticamente con media cero y varianza de σ^2.

a) Demuestre que el estimador de mínimos cuadrador ordinarios (MCO) viene dado por:

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y$$

b) Demuestre que:

$$var(\hat{\beta}_{MCO}) = \sigma^2(X'X)^{-1}$$

Considere ahora un estimador alternativo, el cual se obtiene mediante minimizar la siguiente función objetivo:

$$S(\beta) = (Y - X\beta)'(Y - X\beta) + \lambda(\beta'\beta)$$

donde λ es un escalar no negativo.[78]

c) Encuentre el estimador que minimiza la función de $S(\beta)$ y denomínelo $\hat{\beta}_R$.

d) Muestre que el estimador de $\hat{\beta}_R$ es sesgado a menos que $\lambda = 0$.

e) Encuentre la varianza de $\hat{\beta}_R$.

f) Demuestre que:

$$var(\hat{\beta}_R) \leq var(\hat{\beta}_{MCO})$$

g) ¿Viola el resultado anterior el Teorema de Gauss-Markov?

12. Suponga que se quiere estimar el siguiente modelo en desvíos con respecto a la media:

$$log(peso_i) = \beta_1 cig_i + \epsilon_i$$
$$i = 1, \ldots, N$$

donde *peso* corresponde al peso de un bebé al nacer, y *cig* corresponde al número de cigarrillos que fuma la madre cada día. Por otro lado, ϵ_i contiene todas las variables no observables, en particular, las preferencias de la madre por salud (*salud*), es decir, cuanto la madre valora el estar saludable.

a) Muestre cual es el sesgo que se genera en el estimado de β_1 debido a la omisión de la variable salud antes mencionada.

b) ¿El sesgo es positivo o negativo?

[78]Note que si $\lambda = 0$, entonces, la función $S(\beta)$ corresponde a la función objetivo que minimiza el estimador de MCO.

13. Un investigador quiere estimar la siguiente relación:

$$S_i = \alpha + \beta S_i^P + u_i$$

donde S representa los años de escolaridad del individuo, S^P representan los años de escolaridad (promedio) de los padres del individuo y u es un término de error bien comportado. En otras palabras, se está modelando el impacto que tiene la escolaridad de los padres sobre los niveles de escolaridad de sus hijos. Sin embaro, este investigador ha olvidado incluir la habilidad de los hijos como un determinante clave de su nivel de escolaridad. En realidad, la verdadera relación existente entre la escolaridad de los hijos y la de sus padres debería ser estimada a través de la siguiente ecuación:

$$S_i = \alpha + \beta S_i^P + \gamma H_i + u_i$$

a) Demuestre que el estimador de MCO de β de la primera ecuación está sesgado.

b) ¿De qué depende el sesgo de la estimación?

c) ¿Es posible especular respecto de la dirección de este sesgo?

14. Considere el siguiente modelo que relaciona los salarios con los años de escolaridad de los individuos:

$$y_i = \beta_1 + \beta_2 x_i^* + u_i$$

$$u_i \sim iid(0, \sigma_\mu^2)$$

donde y_i es el logaritmo del salario por hora, x_i^* son los años de escolaridad y u_i es un shock estocástico que se distribuye independiente e idénticamente (idd) con media cero y varianza σ_u^2.

Asuma que la varianza de la variable medida sin error viene dada por $\sigma_{x^*}^2$. Lamentablemente la variable años de escolaridad se encuentra medida con error, por lo tanto, lo que el investigador observa es lo siguiente:

$$x_i = x_i^* + \upsilon_i$$
$$\upsilon_i \sim iid(0, \sigma_\upsilon^2)$$

a) Plantee el modelo que está en condiciones de estimar el investigador.

b) Muestre que la covarianza entre la variable años de escolaridad y el término de error de ese modelo es distinto de cero. ¿Qué consecuencias tiene aquello para el estimador de minímos cuadrados ordinarios (MCO)?

c) ¿Cómo podría obtener un estimador consistente del impacto de los años de escolaridad sobre los salarios?

d) Demuestre que:

$$plim(\overline{\beta}_2) = \beta_2 \left(\frac{\sigma^2_{x^*}}{\sigma^2_{x^*} + \left(\frac{\sigma^2_e}{m}\right)} \right)$$

Ayuda: La expresión $plim(\overline{\beta}_2)$ puede calcularse como:

$$plim(\overline{\beta}_2) = \frac{cov(w, y)}{var(w)}$$

e) ¿Qué sucede con el sesgo a medida que se tiene más mediciones (m) de la variable medida con error?

15. Suponga que un analista desea modelar el puntaje SIMCE obtenido por un alumno de cuarto ciclo básico mediante la siguiente especificación:

$$MAT_i = \alpha + \beta EDM_i + \gamma GEN_i + \tau PEER_i + u_i$$

donde EDM corresponde a los años de escolaridad de la madre del niño, GEN corresponde a una variable dicotómica que tomar el valor 1 si el individuo es mujer y 0 si es hombre, y $PEER$ corresponde a una variable que mide el rendimiento de los compañeros de clases para el alumno i-ésimo. El analista tiene dudas respecto de si la variable $PEER$ es relevante para explicar las variaciones del rendimiento educacional. Por ello, decide recorrer dos regresiones lineales por máxima verosimilitud en base a la ecauación anterior asumiendo que el error aleatorio posee una distribución normal, con media 0 y varianza σ^2, cuyos resultados se muestran a continuación:

Regresores	**Variable dependiente: puntaje Simce Matemáticas**	
	(1)	**(2)**
$\hat{\alpha}$	188,89 (0,41)	147,39 (0,56)
$\hat{\beta}$	5,77 (0,03)	3,16 (0,04)
$\hat{\gamma}$	-3,19 (0,23)	-3,57 (0,23)
$\hat{\tau}$		6,43 (0,06)
LL	-1.028.954,70	-1.023.746,49
SCE	501.855.445,00	475.385.483,00
Observaciones	192.234	192.234

En la tabla el término LL corresponde a la función log-likelihood, evaluada en los estimadores de máxima verosimilitud, y SCE corresponde al cuadrado de los errores estimados, es decir, $l(\hat{\beta}, \hat{\sigma}^2|Y, X)$ y $\sum \hat{u}_i^2$, respectivamente.

a) Interprete los resultados obtenidos en la columna (2) respecto de la variable $PEER$. En base a un test-t, ¿Es $PEER$ una variable relevante?

b) Con la información que se proporciona en el enunciado, realice un test de razón de verosimilitud (LRT) y un test de multiplicadores de lagrange (LM) para testear la hipótesis nula de que τ es igual a 0. ¿Qué concluye?

c) En el caso de que los errores no se distribuyan normales, ¿Cuál de los tres test miraría usted para comprobar si efecticamente $PEER$ es una variable relevante?

d) Si $u_i \sim N(0, \sigma^2)$, la función de densidad de probabilidad de la variable u viene dada por:

$$f(\mu_i|0, \sigma^2) = f(y_i|x_i, \beta, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{u_i}{2\sigma^2}\right)$$

Encuentre, a partir de la expresión anterior, la función de log-likelihood del modelo lineal de regresión.

e) Con lo anterior, encuentre el estimador de máxima verosimilitud (MV) de β y demuestre que los coeficientes estimados por MV coinciden con MCO, es decir: $\hat{\beta}_{MV} = \hat{\beta}_{MCO}$. ¿Qué supuesto es necesario apra que ello se cumpla?

16. Considere el siguiente modelo:

$$y_t = \beta x_t + u_y$$
$$u_t \sim (0, kx_t^2)$$

Además, se tiene que:

$$E(u_t, u_s) = 0$$

$$\forall t \neq s$$

Por otro lado, se dispone de cinco observaciones:

t	1	2	3	4	5
x	1	2	4	1	1
y	2	3	10	1	3

Encuentre el estimador eficiente de β y su varianza.

17. El profesor E.Z Stuff se ha aburrido de los estimadores ed mínimos cuadrados ya que considera que su cálculo es demasiado complicado. Teniendo en cuenta que dos puntos determinan una línea, el profesor Stuff escoge dos observaciones de una muestra de tamaño n y dibuja una línea entre ambas. Luego, él propone la pendiente de esta línea como un estimador alternativo del coeficiente de la

pendiente en el modelo de regresión lineal simple. Algebraicamente, si los dos puntos son (x_1, y_1) y (x_2, y_2) el estimador alternativo del profesor se calcula como:

$$\tilde{\beta}_2 = \frac{y_2 - y_1}{x_2 - x_1}$$

Suponiendo que se cumplen todos los supuestos del modelo de regresión lineal simple,

a) ¿Es $\tilde{\beta}_2$ un estimador lineal?

b) Demuestre que $\tilde{\beta}_2$ es insesgado.

c) Encuentre la varianza de $\tilde{\beta}_2$.

d) Determine la distribución de probabilidades de $\tilde{\beta}_2$

e) ¿Cómo podría convencer al viejo profesor que su estimador no es tan bueno como el de mínimos cuadrados? Explique.

18. El buen profesor E.Z Stuff sigue aburrido de los estimadores de mínimos cuadrados, ya que sigue considerando que su cálculo es demasiado complejo. Insiste, tal como en la prueba anterior, que teniendo en cuenta que dos puntos determinar una linea, se pueden escoger dos obervaciones cualquiera de la muestra y que la pendiente de la línea que une esos puntos es un estimador alternativo del coeficiente de pendiente MC0 en el modelo de regresión lineal simple.

    Algeabraicamente, si los dos puntos son (x_1, y_1) y (x_2, y_2), el estimador alternativo del profesor Stuff es:

$$\tilde{\beta}_2 = \frac{y_2 - y_1}{x_2 - x_1}$$

donde el modelo a estimar viene dado por:

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

El profesor ya esta convencido, gracias a usted, que su estimador es insesgado, lineal y menos eficiente que el estimador MCO. Sin embaro, tiene el presentimiento que su estimador se ve menos afectado por la presencia de hetroscedasticidad que el estimador MCO. En particular, él cree que cuando la varianza del término de error es proporcional al cuadrado de la variable explicativa, su estimador tiene menor varianza que el estimador de MCO. El profesor Stuff no sabe cómo corroborar sus resultados, puesto que no sabe derivar la varianza de $\tilde{\beta}_2$ en presencia de heterocedasticiad. Dada su gran ayuda en la resolución de sus disquisiciones econométricas anteriores, él le solicita que lo ayude a implementar un experimento de Monte Carlo para verificar si efectivamente su estimador es más efciente que el de MCO cuando existe heteroscedasticidad. Para ello, suponga que:

$$(\beta_1, \beta_2)' = (2, 3)'$$
$$E(u_i|x_i) = 0$$
$$E(u_i^2|x_i) = \sigma^2 x_i^2$$
$$\sigma^2 = 4$$

Se pide:

a) Explique detalladamente cómo puede llevar a cabo el experimento de Monte Carlo en este caso. Suponga que el tamaño de muestra en el que está inicialmente interesado es n=100. Debe explicar cómo generaría la variable independiente (x), el término de error y la variable dependiente (y). (Ayuda: suponga que el programa que utiliza para realizar su experimiento posee un generador de números aleatorios que sólo genera números aleatorios provenientes de una normal estándar).

b) ¿Cómo puede verificar, en base a sus resultados de la pregunta anterior, qué estimador es efectivamente más eficiente en presencia de heteroscedasticidad? Sea específico en su respuesta.

c) ¿Podría usted verificar de alguna manera si ambos estimadores son igualmente eficientes asintóticamente? ¿Cómo lo haría? Explique.

19. Considere el modelo de regresión presentado en la siguiente tabla, donde se exploran los determinantes del salario para un grupo de 935 trabajadores. La variable dependiente es el logaritmo natural del salario; *educ*, *exper* y *tenure*, corresponde, respectivamente, a la educación, experiencia y antigüedad en el empleo actual, *married* y *black* son dos variables dummies que toman el valor uno si el individuo es, respectivamente, casado y de raza negra; IQ es el coeficiente intelectual; *age* es la edad del individuo, *urban* es una variable dummy que toma el valor de uno si el individuo provene de una zona urbana; y $educ2$ y $exper2$ corresponde al cuadrado de la variable educación y experiencia, respectivamente. También se presenta la matriz de varianzas y covarianzas de los coeficientes estimados de la regresión:

Dependent variable: lwage

	Coefficient	*Std.Error*	*t-rartio*	*p-value*	
Const	4.11152	0.548232	7.4996	<0.00001	***
educ	0.151054	0.0741564	2.0370	0.04194	**
exper	0.0224518	0.0127881	1.7557	0.07947	*
tenure	0.0111113	0.00247841	4.4832	<0.00001	***
married	0.192796	0.0389981	4.9437	<0.00001	***
black	-0.160094	0.0390209	-4.1028	0.00004	***
IQ	0.00393348	0.000997834	3.9420	0.00009	***
age	0.0106494	0.00499086	2.1338	0.03312	**
urban	0.190289	0.0266745	7.1337	<0.00001	***
$educ^2$	-0.00358143	0.00259694	-1.3791	0.16820	
$exper^2$	-0.000516399	0.000562781	-0.9176	0.35908	

Mean dependt var	6.779004	S.D dependent var	0.421144
Sums squared resid	122.5440	S.E of regression	0.364175
R-squared	0.260251	Adjusted R-squared	0.252245
F(10,924)	35.50728	P-value(F)	3.00e-54
Log-likelihood	-376.7119	Akaike criterion	775.4239
Schwarz criterion	828.6699	Hannan-Quinn	795.7270

Const	Educ	Exper	tenure	married	
0,300558	-0,0379323	-8,59E-04	4,14E-05	0,00103717	const
	0,00549917	-5,26E-05	-8,78E-07	-2,08E-05	educ
		1,64E-04	1,08E-07	1,53E-05	exper
			6,14E-06	-3,09E-06	tenure
				1,52E-03	married

black	IQ	age	urban	educ2	exper2	
-0,00203152	-1,54E-05	-9,20E-04	9,50E-05	1,33E-03	3,37E-05	const
7,59E-06	-1,12E-05	2,37E-05	-6,13E-05	-1,92E-04	3,77E-06	educ
2,06E-05	3,03E-07	1,12E-05	6,20E-07	1,66E-06	-6,89E-06	exper
4,86E-06	-1,39E-07	-1,70E-06	1,89E-06	2,53E-08	-5,90E-08	tenure
8,28E-05	1,33E-07	-8,60E-06	3,05E-05	9,81E-07	-9,15E-07	married
1,52E-03	1,35E-05	1,05E-05	-9,16E-05	-5,66E-07	-1,05E-06	black
	9,96E-07	3,86E-07	-8,82E-07	2,79E-07	-2,12E-08	IQ
		2,49E-05	-2,36E-06	-1,24E-06	-9,19E-07	age
			7,12E-04	1,87E-06	1,60E-08	urban
				6,74E-06	-1,04E-07	educ2
					3,17E-07	exper2

De los resultados de la estimación de MCO recién presentados se observa que los individuos casados tienden a tener un mayor salario que los solteros, mientras que los individuos de color tienden a ganar menos que los blancos.

a) ¿Cómo testaría la hipótesis que raza y estado marital se compensan, en el sentido que los individuos casados y blancos ganan lo mismo que los solteros y negros? Con la información entregada, ¿puede hacerlo? Si puede, hágalo, especificando la forma del test, su distribución bajo la nula, grados de libertad y el criterio de norechazo o rechazo de la misma.

b) ¿Qué puede decir de la existencia de no linealidades en el modelo? Explique su respuesta.

Considere ahora el proceso iterativo de eliminación de variables explicativas basados en eliminar, una a una consecutivamente, la variable con el test t más pequeño en el valor absoluto, hasta llegar a un modelo con una sola variable explicativa (los resultados de este proceso se muestran al final). En base a estos resultados:

c) ¿Cuál sería el modelo con el que usted se quedaría? ¿Cómo lo determinaría? Explique en base a qué criterios(s) realizaría su selección de modelo. Realice los tests formales necesarios para apoyar su respuestas.

20. Utilizando los datos de la encuesta de Caracterización Socioeconómica Nacional (CASEN) 2009 se ha estimado el siguiente modelo para explicar los salarios de los individuos:

$$y_i = \beta_1 + \beta_2 S_i + \beta_3 X_i + \beta_4 X_i^2 + \beta_5 M_i + \beta_6 I_i + \beta_7 M_i I_i + \mu_i$$

donde y es el logaritmos del salarios por hora, S son los años de escolaridad, X son los años de experiencia laboral, M es una variable dummy que toma el valor de 1 si el individuo es mujer, I es una variable dummy que toma el valor de 1 si el individuo es indígena, y u es un shock estocástico que se distribuye independiente e idénticamente con media cero y varianza σ^2.

Los resultados de la estimación econométrica de este modelo se muestran a continuación, así como también la matriz de vairanza y covarianzas de los coeficientes estimados:

	Coeficiente
Constante	5,515385
S	0,1292775
X	0,0146205
X^2	-0,0000176
M	-0,1708808
I	0,0845115
MI	0,0067471
Número de observaciones	54136
Suma de los errores al cuadrado	161.141.141
Suma de los cuadrados del modelo	261.260.501

	S	X	X^2	M	I	MI	Constante
S	3.945e-06						
X	4.975e-08	1.431e-06					
X^2	8.347e-09	-2.713e-08	5.838e-10				
M	-6.732e-06	-1.328e-06	2.302e-08	1.189e-04			
I	2.378e-06	3.051e-08	1.83e-08	3.524e-05	3.8379e-04		
MI	-9.061e-08	1.739e-06	-4.252e-08	-1.0719e-04	-3.8278e-04	1.32127e-03	
Constante	-4.55e-05	-1.364e-05	1.16e-07	4.442e-05	-7.103e-05	2.894e-05	7.0159e-04

Nota: El numero 3.945e-06 equivale a 0.000003945. La matriz de varianza y covarianza se obtuvo por medio de la corrección de White.

a) Al 5 por ciento de significancia, estableza si las variables S, X, M e I son estadísticamente significativas.

b) Al 5 por ciento de significancia, ¿es el modelo globalmente significativo?

c) Calcule el R^2 y el $\bar{R}^2$.

d) ¿Cuál es el efecto que tiene un año adicional de experiencia laboral en los salarios? Haga un gráfico para explicar su respuesta.

e) ¿Cuál es la diferencia salarial entre las mujeres indígenas y las mujeres no indígenas?

f) ¿Cuál es la diferencia salarial entre homobres indígenas y las mujeres no indígenas?

g) Evalúe la siguiente hipótesis nula al 5 por ciento:

$$H_0 : \beta_2 + \beta_5 = 0$$

h) ¿Qué consecuencias tendría el hecho de que la variable años de escolaridad estuviera medida con error? Por ejemplo, qué ocurriría si:

$$S = S^* + v$$

donde S^* es la variable años de escolaridad sin error de medida, y v es un shock estocástico *idd* con media cero y varianza homocedástica.

i) ¿Cómo solucionaría el problema anterior? ¿Qué método de estimación emplearía? Explique.

21. Utilizando datos para EE.UU se ha estimado el siguiente modelo de participación laboral para las mujeres:

Variables independientes	MPL (MCO)	Probit
nwifeinc	-0.0034 (0.0015)	-0.012 (0.005)
educ	0.038 (0.007)	0.131 (0.025)
exper	0.039 (0.006)	0.123 (0.019)
exper^2	-0.0006 (0.00018)	-0.0019 (0.0006)
age	-0.016 (0.002)	-0.053 (0.008)
kidslt6	-0.262 (0.032)	-0.868 (0.119)
kidse6	0.013 (0.013)	0.036 (0.043)
constante	0.586 (0.151)	0.27 (0.509)
R^2	0.264	0.221

La variable dependiente es una variable dummy que toma el valor 1 si la persona participa en el mercado laboral (ocupado o desocupado) y cero si no (inactivo). Las variables explicativas con el ingreso del cónyuge (nwifeinc) en miles de dólares, años de escolaridad (educ), experiencia laboral (exper) y su cuadrado, edad (age), número de hijos menores de seis años (kidslt6), y número de hijos entre 6 y 18 años (kidse6). En base a esta información, responda lo siguiente:

a) ¿Qué diferencia conceptual existe entre los modelos MPL y Probit? (Ayuda: Refiérase al método de estimación y los efectos marginales).

b) En el MPL, ¿cuál es el efecto que tiene un niño menor de seis años en la probabilidad de participar en el mercado laboral de una mujer?

c) En el modelo Probit para una mujer que no tiene niños menores de 6 años ¿Cuál sería el efecto que tendría la probabildiad de participar en el mercado laboral pasar de kidslt6=0 a kidslt6=1? Los valores promedios muestrales para el resto de las variables son los siguientes: age=42.5, educ=12.3, exper=10.6, kidsge6=1 y nwifeinc=20.13.

d) Calcule ahora cúal sería el efecto de pasar de kidslt6=1 a kidslt6=2. ¿Cómo se compara con el efecto anterior?

22. Se propone el siguiente modelo para el ahorro (Y) de EE.UU para el periodo 1970-1995 (en millones de dólares):

$$Y_t = \beta_1\beta_2 DM_t + \beta_3 X_t + \beta_4 DMX_t + u_t$$

donde X es el ingreso (en millones de dolares), DM es una variable dummy que toma el valor 1 para los años comprendidos entre 1970 y 1981 (ambos inclusive), y cero si no. Basándose en los resultados que se presentan a continuación responda las siguientes preguntas:

a) Estime el ahorro medio de los EE.UU en el año 1971, si en dicho período se tuvo un ingreso de US790.2 (millones de dolares).

b) ¿Es posible afirmar que existe una diferencia estructural en la regresión entre los periodos 1970-1981 y 1982-1995? Justifique su respuesta.

Dependent Variable: Ahorros
Method: Leas Squares
Sample: 1970 1995
Included observations: 26

Variable	**Coefficient**	**Std.Error**	**t-Statistic**	**Prob.**
C	1.016117	20.16483	0.050391	0.9603
DM	152.4786	33.08237	4.609058	0.001
INGRESO	0.080332	0.014497	5.541347	0.0000
DMX	-0.065469	0.015982	-4.096340	0.0005
R-squared	0.881944	Mean dependent var		162.0885
Adjusted R-squared	0.865846	S.D dependent var		63.20446
S.E of regression	23.14996	Akaike info criterion		9.262501
Sum squared resid	11790.25	Schwarz criterion		9.456055
Log likelihood	-116.4125	F-statistic		54.78413
Durbin-Watson stat	1.648454	Prob(F-statistic)		0.000000

23. Suponga que está interesado en estimar el ingreso por ventas de una tienda, para lo cual dispone de 90 datos sobre las siguientes variables: ingreso total en millones de pesos (y), el número de competidores en el mercado (nc), el gasto en publicidad en millonres de pesos (gp) y el número de vendedores (nv). Usted debe estimar el siguiente modelo de regresión lineal:

$$y_i = \beta_0 + \beta_1 nc_i + \beta_2 gp_i + \beta_3 nv_i + u_i$$

donde las matrices $(X'X)^{-1}$ y $X'Y$ vienen dadas por:

$$(X'X)^{-1} = \begin{pmatrix} 5 & -3 & 2 & 0 \\ -3 & 6 & -2 & -4 \\ 2 & -2 & 4 & 3 \\ 0 & -4 & 3 & 4 \end{pmatrix}$$

$$X'Y = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 2 \end{pmatrix}$$

y además:

$$\hat{\sigma}^2 = 0,5$$

$$\sum_{i=1}^{90} (Y_i - \overline{Y})^2 = 80$$

El término de error (u) corresponde a un ruido blaco con media cero y varianza constante (σ^2).

a) Encuentre el estimador de mínimos cuadrador ordinarios (MCO) de los parámetros de este modelo.

b) Encuentre la matriz de varianzas y covarianzas del estimador MCO.

c) ¿Son estadísticamente significativos (individualmente) los coeficientes estimados?

d) Determine la bondad de ajuste del modelo a través del R^2 y del R^2 ajustado. ¿Qué medida es más confiable? Explique.

e) Testee la hipótesis de que todos los coeficientes, salvo la constante, son cero (significancia global del modelo).

f) Construya un intervalo de confianza al 95 por ciento para β_2 y σ^2.

24. Sea X una variable que se distribuye normal con media μ y varianza σ^2. Se han obtenido dos muestras aleatorias de X de tamaños T_1 y T_2 respectivamente, con medias muestrales $\overline{X}_1$ y $\overline{X}_2$.

    Un investigador propone dos estimadores alternativos de μ:

$$\hat{\mu}_1 = \frac{\overline{X}_1 + \overline{X}_2}{2}$$

$$\hat{\mu}_2 = \frac{T_1\overline{X}_1 + T_2\overline{X}_2}{T_1 + T_2}$$

a) ¿Son insesgados estos estimadores?

b) ¿Cuál estimador es más eficiente?

Posteriormente, se busca estimar μ^2 y se proponen 3 estimadores:

$$\hat{\mu}_1^2 = \overline{X}_1\overline{X}_2$$

$$\hat{\mu}_2^2 = \left(\frac{\overline{X}_1 + \overline{X}_2}{2}\right)^2$$

$$\hat{\mu}_3^2 = \frac{\overline{X}_1^2 + \overline{X}_2^2}{2}$$

a) ¿Cuáles son los sesgos de cada uno de los estimadores?

b) Solo considerando el criterio de insesgamiento, ¿cuál estimador es mejor?

25. Un investigador dispone de una muestra de 50 países, con datos sobre el número de diarios comprados por un adulto al año (Y), y sobre el producto interno bruto per cápita (X), medido en pesos. Con estos datos, el investigador estima el siguiente modelo:

$$\hat{Y} = \underbrace{25}_{10} + \underbrace{0.020}_{0.010} X$$

se sabe que $R^2 = 0.06$, $SCR = 4000$, y $F = 4$. El valor que aparece debajo del coeficiente muestra el error estándar de los estimadores, SCR es la suma de los cuadrados de los residuos, y F es el test F de significancia global de la regresión.

Muestre, siendo riguroso matemáticamente, cómo los siguiente componentes asociados a la regresión cambian si el producto interno bruto (X) en vez de estar medido en pesos, estuviera medido en dólares (asuma que mantiene la siguiente relación peso-dólar: 1 peso = 2 dólares).

26. Se recopiló información sobre características de los hogares de 1884 niños que rindieron la prueba SIMCE de cuarto básico el año 2005, en la región de la Araucanía y sólo para establecimientos rurales. En base a dicha información se estimó el siguiente modelo:

$$MATE_i = \beta_1 + \beta_2 EDM_i + \beta_3 EDP_i + \beta_4 RC_i + \beta_5 YPC_i + u_i$$

en donde:

$MATE$: es el puntaje SIMCE para la prueba de matemáticas.

EDM: son los años de educación de la madre del niño.

EDP: son los años de educación del padre del niño.

RC: es un indicador de recursos educativos en la casa (mide disponibilidad de recursos para el estudio en casa, tales como diccionarios, enciclopedias, espacio adecuado para el estudio, etc).

YPC: corresponde al ingreso per cápita del hogar (en decenas de miles de pesos).

La estimación por MCO arrojó los siguiente resultados (las desviaciones estándar de los parámetros estimados se encuentre entre paréntesis):

$$\begin{array}{lccccc} \widehat{MATE}_i = & 169{,}61 & +\ 2{,}43\ EDM_i & +\ 2{,}24\ EDP_i & +\ 20{,}74\ RC_i & +\ 1{,}07\ YPC_i \\ & (3{,}15) & (0{,}41) & (0{,}40) & (4{,}76) & (0{,}41) \end{array}$$

$$R^2 = 12,8$$

$$\widehat{Cov}(\hat{\beta}_2, \hat{\beta}_3) = -0,07$$

En base a esta información se pide lo siguiente:

a) Interprete el resultado de estimación. prese atención a los signos de los coeficientes y en la significancia estadística individual de cada uno de ellos.

b) Pruebe la siguiente hipótesis nula $H_0 : \beta_2 = \beta 3 = \beta 4 = \beta 5 = 0$. Interprete.

c) Muestre cómo podría comprobar la hipótesis de que los años de educación del padre y de la madre influyen de manera equivalente en el puntaje SIMCE del niño. Explique el estadístico que usará, su distribución, y luego calcule e interprete los resultados.

27. Considere el siguiente modelo de regresión lineal simple:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

donde:

$$f(u_i) = \frac{1}{\lambda} e^{\left(-\frac{u_i}{\lambda}\right)}$$

$$u_i \geq 0$$

Note que en este modelo los errores son positivos, de hecho:

$$E(u_i) = \lambda \tag{61}$$

a) Demuestre que el estimador MCO de β_2 es insesgado.

b) Dmuestre que el estimador de β_1? es sesgado.

c) ¿Es consistente el estimador MCO de β_1?

d) Demuestre que $E(u_i) = \lambda$.

28. Considere el siguiente modelo:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

donde el shock estocástico es independiente e idénticamente distribuido con:

$$u_{\sim} N(0, \sigma^2)$$

En base a esta información:

a) Demuestre formalmente que los estimadores de mínimos cuadrados ordinarios (MCO) de β_1 y β_2 vienen dados por:

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 X_i + \mu_i$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

b) Encuentre (i) $var(\hat{\beta}_1)$, (ii) $var(\hat{\beta}_2)$, (iii) $cov(\hat{\beta}_1, \hat{\beta}_2)$.

c) Suponga que usted cuenta con los siguiente datos:

X	1	2	4	3
Y	-1	0	2	3

Calcule el estimador de MCO de β_1 y β_2.

d) Calcule el estimador insesgado de la varianza del error ($\tilde{\sigma}^2$).

e) ¿Es $\hat{\beta}_2$ estadísticamente significativo?

f) Construya un intervalo de confianza al 95 por ciento para β_1.

g) Calcule el R^2 y el $\overline{R}^2$. Si tuviera que evaluar desde un punto de vista estadístico la inclusión de un segundo regresor Z, ¿qué estadístico consideraría? Explique su respuesta.

h) Sabiendo que la varianza del error de una predicción puntual viene dado por:

$$var(e_0) = \tilde{\sigma}^2[1 + x_0'(X'X)^{-1}x_0]$$

en donde $x_0' = (1\ X_0)$, realice una predicción puntual para $X_0 = 5$ y construya un intervalo de confianza al 95 por ciento para ella.

i) Un supuesto clave para hacer inferencia estadística es que el término de error (u) se distribuya normal. Para verificar aquello realice un test de Jarque-Bera.

29. Considere la siguiente regresión utilizada para evaluar el efecto de la educación y la experiencia sobre el salario de un individuo:

$$ln(w_i) = \beta_1 + \beta_2 Educ_i + \beta_3 Exper_i + \beta_4 IQ_i + \epsilon_i$$

La variable dependiente corresponde al logaritmo natural del salario, la cual es explicada a través del nivel educacional ($Educ$) del individuo i, de su experiencia laboral ($Exper$) y de su coeficiente intelectual (IQ). Los resultados de estimar este modelo para una muestra (n) de 935 individuos se presenta en la siguiente tabla:

Tabla 1: MCO, using observations 1-935
Variable dependiente: ln(wi)

	Coefficient	*Std.Error*	*t-ratio*	*p-value*	
Constant	5.19808	0.121543	427676	<0.00001	***
IQ	0.00578564	0.0000979658	5.9058	<0.00001	***
Educ	0.057108	0.00734796	7.7720	<0.00001	***
Exper	0.0195249	0.00324435	6.0181	<0.00001	***

Mean dependent var	6.7799004	S.D dependent var	0.421144
Sum squared resid	138.7795	S.E of regression	0.386089
R-squared	¿?	Adjusted R-squared	¿?
F(3,931)	¿?	P-value(F)	¿?
Log likelihood	-434.8765	Akaike criterion	877.7529
Schwarz criterion	897.1151	Hannan-Quinn	8551359

Tabla 2: Matriz de Varianzas

Const	IQ	educ	exper	
0.0147726	-5.05316E-05	-0.00049199	-0.000248094	const
	9.59730E-07	-3.42943E-06	-4.17775E-08	IQ
		5.399250E-07	9.697330E-06	educ
			1.05258E-05	exper

Tabla 3: Estadística Descriptiva

Variable	**Mean**	**Median**	**Minimum**	**Maximum**
IQ	101.282	102.0000	50.000	145.000
educ	13.4684	12.0000	9.00000	18.0000
exper	11.5636	11.0000	1.00000	23.00000
Variable	**Std. Dev**	**C.V.**	**Skewness**	**Ex.Kurtosis**
IQ	15.0526	0.148621	-0.340425	-0.022965
educ	2.19665	0.163096	0.547796	-0.737349
exper	4.37459	0.378305	0.077676	0.567195

a) ¿Tienen los coeficiente obtenidos los signos esperados? Explique.

b) Muestre que la función de regresión muestral para por las medias muestrales.

c) Realice los test de significancia individual para cada una de las variables independientes (incluya la constante). Debe especificar el test a utilizar, sus grados de libertad, la hipótesis nula y los criterios de rechazo o de no rechazo.

d) Obtenga el R^2 y el R^2 ajustado de la regresión.

e) ¿Cómo haría un test de significancia conjunta? ¿Puede hacerlo? Si puede, hágalo, especificando el test a utilizar, sus grados de libertad, la hipótesis nula y los criterios de rechazo y de no rechazo. Si no puede, explique qué información adicional requeriría.

Suponga ahora que desea testar la siguiente proposición: "Un año de educación es más valioso, en términos de generación de renta, que un año de experiencia laboral".

f) Establezca la hipótesis nula para testar esta proposición. Debe especificar el test a utilizar, sus grados de libertad, la hipótesis nula y los criterios de rechazo y de no rechazo.

g) Con los datos presentados, ¿tiene suficiente información para hacer el test? Si puede hágalo.

h) ¿Cuál sería el salario (no su logaritmo) estimado para un individuo con un coeficiente intelectual de 100, 12 años de educación y sin experiencia laboral? Explique.

i) ¿Cómo testaría la hipótesis en (f) utilizando un test F? Explique detalladamente su procedimiento.

30. Considere los siguientes datos:

i	**X**	**Y**
1	1	2
2	3	5
3	4	7
4	6	9

Por lo tanto, el modelo a estimar presenta la siguiente estructura:

$$Y_i = \beta_0 + \beta_1 X_1 + u_i$$

donde u representa el término de error, el cual cumple con todos los supuestos estándares de mínimos cuadrados ordinarios (MCO).

a) Encuentre los estimadores de MCO para β_0 y β_1 (trabaje en forma matricial sus resultados. Utilice sólo tres decimales. Sea explícito en la manera en que invierte sus matrices).

b) Encuentre el estimador de la varianza del término de error (σ^2). Note que está trabajando con un número pequeño de observaciones.

c) Encuentre la matriz de varianzas y covarianzas de los estimadores de MCO.

d) Encuentre el R^2 y R^2 ajustado. ¿En cuál se fijaría usted? ¿Por qué?

e) ¿Son estadísticamente significativos los estimadores MCO para β_0 y β_1?.

f) Si su estimación de MCO para β_1 no fuera estadísticamente significativa, ¿significa esto que debería eliminar este regresor y reemplazarlo por otro? Explique.

g) Construya los intervalos de confianza al 95 para los estimadores MCO de β_0 y β_1.

31. Considere el siguiente modelo:

$$Y_i = \beta + u_i$$

donde $i = 1, \ldots, n$, y u_i es un término de error bien comportado, eso es, se distribuye de manera independiente, con media cero y varianza constante (σ^2).

a) Demuestre formalmente que el estimador de mínimos cuadrados ordinarios (MCO) para β viene dado por:

$$\hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} Y_i = \overline{Y}$$

b) Encuentre el valor esperado, la varianza y el error cuadrático medio (ECM) de este estimador.

c) A partir de los siguientes datos encuentre $\hat{\beta}_{MCO} : Y_i = \{2, 3, 5, 5, 7, 8\}$.

d) Considere el siguiente estimador alternativo:

$$\tilde{\beta} = \frac{n\overline{Y}}{n+1}$$

Encuentre el valor esperado, la varianza y el ECM de este estimador.

e) Utilizando la misma muestra de datos anteriores encuentre $\tilde{\beta}$.

f) ¿Cuál de estos dos estimadores prefiere usted? Justifique muy bien su respuesta.

32. Para el caso del siguiente modelo simple:

$$Y_i = a + bX_i + \epsilon_i$$

donde el término de error cumple con los supuestos habituales:

a) ¿Cuál es el criterio para encontrar el estimador de mínimos cuadrados ordinarios (MCO)?

b) Encuentre las ecuaciones normales de este problema.

c) Encuentre los estimadores de MCO para a y b.

d) Encuentre las varianzas de los estimadores de MCO de a y b.

33. Suponga que Pepe quiere estimar por MCO el siguiente modelo:

$$Y_i = \alpha + u_i$$

A Pepe le entró un virus en su computador y por lo tanto no le fue posible estimar usando su software econométrico. Decide, por lo tanto calcular "a manos" el modelo. Los resultados que él obtiene indican que $\hat{\alpha} = 3.5$ con una desviación estándar de 0.1 , el test t asociado comprueba que el parámetro estimado es significativamente distinto de 0 y el R^2 lo calcula en 0.86. Dado los resultados obtenidos, afirma Pepe, el modelo es una buena representación de los datos. ¿Es correcto el cálculo que hizo Pepe?

34. Suponga que se tiene las siguientes dos regresiones para una muestra de 1000 individuos:

$$Log(\widehat{salario_i}) = 12 + 0,07esc_i + 0,02exper_i + 0,002capitalsocial_i$$

$$SCE^1 = 15.000$$

$$Log(\widehat{salario_i}) = 12 + 0,08esc_i + 0,03exper_i$$

$$SCE^1 = 18.000$$

donde *esc* corresponde a los años de escolaridad, *exper* a los años de experiencia, *capitalsocial* a una variable que aproxima la cantidad de redes sociales a la que pertenece el individuo y SCE a la sumatoria de los errores estimados al cuadrado. ¿Es recomendable eliminar la variable *capitalsocial*? ¿Es posible construir un test t asociado al parámetro de aquella variable?

35. El terremoto que afectó a la zona centro sur del país en el año 2010 dejó una serie de daños en términos de infraestructura. Para solventar los costos de la reconstrucción el poder ejecutivo ha propuesto una reforma tributaria que busca incrementar de manera significativa la recaudación fiscal. Dado que el solo crecimiento económico genera una mayor recaudación tributaria, un analista propone estimar el siguiente modelo para determinar cuál es la elasticidad recaudación tributaria respecto del crecimiento económico:

$$R_t = \alpha_1 + \beta_2 PIB_t + \alpha_3 IVA_t + u_t$$

con $t = 1990, \ldots, 2009$. R es la recaudación tributaria (medido en miles de millones de dólares), PIB es el producto interno bruto (medido en miles de millones de dólares), IVA es la tasa del impuesto al valor agregado (actualmente esta tasa es de 19 por ciento), y u representa una perturbación esférica. En base a esta información responda las siguientes preguntas:

a) Identifique claramente cuáles son los supuestos que se están haciendo al modelar de esta manera la relación existente entre la recaudación tributaria con el PIB y el IVA.

b) Encuentre el estimador de mínimos cuadrados ordinarios para α_1, α_2 y α_3. Es decir:

$$\hat{\alpha} = (X'X)^{-1}X'R$$

donde X es una matriz de 20x3, y R es un vector de 20x1. Ayuda: no es necesario que invierta la matriz de $(X'X)$, solo se deben indentificar claramente los elementos de esta solución.

Suponga que las estimaciones por MCO arrojan los siguientes resultados:

	Coeficiente	**Error estándar**	**p-value**
$\hat{\alpha}_1$	¿?	0,123	¿?
$\hat{\alpha}_2$	0,07	0,003	¿?
$\hat{\alpha}_3$	0,03	¿?	0,025
R^2	0,456		
$\bar{R}^2$	0,354		

c) Interprete los coeficientes estimados para $\hat{\alpha}_2$ y $\hat{\alpha}_3$.

d) Si se quiere financiar la reconstrucción solo con el crecimiento del PIB, ¿cuál debería ser el PIB del año 2010 si la reconstrucción tiene un costo de 15 mil millones de dólares?

e) ¿Son estadísticamente significativos los estimadores para α_2 y α_3?

f) Suponga que se está analizando la posibilidad de incorporar la tasa de desempleo como variable explicativa. ¿Qué estadístico miraría usted para determinar su incorporación? Justifique su respuesta.

36. En un estudio sobre las ventas de tabaco en EE.UU., se ha especificado el siguiente modelo:

$$ln(V_t) = \beta_1 + \beta_2 ln(P_t) + \beta_3 ln(GC_t) + \beta_4 GP_t + \upsilon_t$$

donde V son las ventas (en millones de cigarrillos) de las principales empresas tabacaleras, P es el precio unitario de los cigarrillos (en dólares de 2008), GC son los gastos en publicidad, en cine, televisión y radio (en miles de dólares de 2008) y GP son los gastos en publicidad en prensa escrita y carteles publicitarios (en miles de dólares de 2008). La variable υ representa una perturbación esférica.

Los resultados de la estimación realizada por mínimos cuadrados ordinarios (MCO) del modelo anterior, usando datos desde 1960 hasta 2008 (ambos inclusive), son los siguientes:

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\beta}_4 \end{pmatrix} = \begin{pmatrix} 2,5 \\ -0,3 \\ 0,04 \\ 1,45 \end{pmatrix}$$

$$var(\hat{\beta}) = \begin{pmatrix} 1,47 & -0,3 & 0,0028 & -0,05 \\ . & 0,068 & -0,004 & 0,003 \\ . & . & 0,002 & -0,003 \\ . & . & . & 0,011 \end{pmatrix}$$

a) Interprete desde un punto de vista económico los resultados de la estimación de este modelo. ¿Tiene sentido que los signos y las magnitudes de los coeficientes estimados?

b) Contraste la hipótesis nula $H_0 : \hat{\beta}_3 = \hat{\beta}_4$ versus la alternativa $H_a : \hat{\beta}_3 \neq \hat{\beta}_4$. Interprete económicamente el resultado de esta prueba.

c) Algunos piensan que una probable medida para controlar las ventas de tabaco consistiría en limitar el volumen de publicidad; sin embargo, otros creen que esta medida no sería eficaz ya que, según ellos, la publicidad no afecta las ventas del tabaco. A la luz de los resultados de la regresión ¿qué opina usted? ¿Cómo testaría formalmente la hipótesis que la publicidad no afecta el consumo de tabaco? ¿Puede realizar el test a partir de la información entregada? Si puede, hágalo.

37. John Taylor (1993), influyente macroeconomista, sugirió que las autoridades monetarias seguían la regla con la cual ajustaban la tasa de interés de política monetaria a cambios en la inflación y en la brecha del producto $(y_t - \overline{y})$. Esta regla puede expresarse de la siguiente manera:

$$i_t = \beta_1 + \beta_2(\pi_t - \overline{\pi}) + \beta_3(y_t - \overline{y})$$

donde i es la tasa de interés, π es la tasa de inflación, $\overline{\pi}$ es la tasa de inflación del largo plazo que el Banco Central define como su objetivo de política monetaria (asuma que dicho objetivo es de 3), y es el producto interno bruto, y finalmente, $\overline{y}$ es el producto de largo plazo. Luego, el modelo a estimar puede plantarse de la siguiente manera:

$$i_t = \beta_1 + \beta_2(\pi_t - \overline{\pi}) + \beta_3(y_t - \overline{y}) + u_t$$

donde u es un shock estocástico que se distribuye independiente e idénticamente (normal) con media cero y varianza homocedástica. Una estimación econométrica por mínimos cuadrados ordinarios (MCO) con datos para *happyland* (20 datos) arrojó los siguientes resultados:

	Coeficiente	**Error estándar**	**Test t**	**p-value**
$\hat{\beta}_1$	3	0,587		
$\hat{\beta}_2$	2	0,65		
$\hat{\beta}_3$	1,5		2,679	0,023

Además, se sabe que $\hat{u'}\hat{u} = 4587$ y la varianza muestras de la variable dependiente es 2480.[79] En base a esta información responda lo siguiente:

a) Explique lo más detalladamente y con el máximo de formalidad posible la manera en que se calculan los coeficientes del modelo por mínimos cuadrados ordinarios ($\hat{\beta}_1$,$\hat{\beta}_2$, $\hat{\beta}_3$). ¿Qué rol juegan las condiciones de primer y segundo orden?

b) Dentro del contexto de este modelo, ¿qué tipo de fenómenos podría estar capturando el shock estocástico u?

c) ¿Son estadísticamente significativos ($\alpha = 5$) los efectos de la brecha del producto y de la brecha de inflación sobre la tasa de interés? Defina claramente la hipótesis nula, la hipótesis alternativa, la distribución del test y los criterios de rechazo.

d) Construya un intervalo de confianza al 95 por ciento para β_2.

e) Evalúe la significancia global del modelo. Defina claramente la hipótesis nula, la hipótesis alternativa, la distribución del test y los criterios de rechazo.

f) Calcule la bondad de ajuste de la regresión (R^2). ¿Qué crítica podría formularle a este estadístico desde el punto de vista conceptual?

g) Calcule ahora el R^2 ajustado ($\overline{R}^2$) de la regresión. ¿Qué ventaja presenta éste por sobre el R^2.

[79] La varianza de la variable dependiente se ha calcula de la siguiente manera:

$$var(i) = \frac{1}{n}\sum_{t=1}^{T}(i_t - \bar{l})^2$$

h) Si para este año se espera que la inflación se ubique en 4 y asumiento que la brecha del producto es cero, ¿qué debíera hacer el Banco Central de acuerdo a la regla de Taylor? Justifique desde un punto de vista económico su respuesta.

38. Considere el siguiente modelo:

$$sleep = \alpha_1 + \alpha_2 totwrk + \alpha_3 educ + \alpha_4 age + u$$

donde *sleep* es el tiempo de sueño (medido en minutos por semana), *totwrk* es el tiempo de trabajo (medido en minutos por semana), *educ* son los años de escolaridad, *age* es la edad de la persona, y u es un shock estocástico independiente e idénticamente distribuido (normal) con media cero y varianza σ^2. La siguiente tabla presenta las estimaciones que se llevaron a cabo por mínimos cuadrados ordinarios:

	Modelo 1	**Modelo 2**
Constante	3638.25 (112.28)	3586.38 (38.91)
totwrk	-0.148 (0.017)	-0.151 (0.017)
educ	-11.13 (5.88)	
age	2.20 (1.45)	
Número de observaciones	706	706
R^2	0.113	0.103

a) En el modelo 1, ¿son *educ* y *age* individualmente significativas al 5 por ciento usando contrastes de dos colas? Explique su respuesta.

b) En el modelo 1, ¿son *educ* y *age* conjuntamente significativas al 5 por ciento? Justifique su respuesta.

c) El incluir *educ* y *age* en el modelo, ¿afecta mucho la disyuntiva estimada entre dormir y trabajar?

39. Marc Nerlove ha estimado la siguiente función de costos para la generación de electricidad:

$$Y = AX^{\beta} P_1^{\alpha_1} P_2^{\alpha_2} P_3^{\alpha_3} u$$

donde Y es el costo total de producción, X es el producto en horas kilovatio, P_1 es el precio del insumo trabajo, P_2 es el precio del insumo capital, P_3 es el precio del combustible y μ es un shock estocástico bien comportado. Teóricamente, se espera que la suma de las elestacidades-precio sea igual a la unidad, es decir:

$$\alpha_1 + \alpha_2 + \alpha_3 = 1$$

Sin embargo, al imponer esta restricción, la función de costos anterior puede escribirse de la siguiente manera:

$$\frac{Y}{P_3} = AX^{\beta}\left(\frac{P_1}{P_3}^{\alpha_1}\right)\left(\frac{P_2}{P_3}\right)^{\alpha_2} u$$

En otras palabras, la primera es una función de costo no restringida y la segunda es una función de costo restringida. Con base en una muestra de 29 empresas de tamaño mediano y después de realizar la transformación logarítmica, Nerlove estimó las siguiente regresiones:

$$ln(Y_i) = -4,93 + 0,94ln(X_i) + 0,31ln(P_1) - 0.26ln(P_2) + 0.44ln(P_3) + \hat{\epsilon}_i$$

$$\text{con } SCE = 0.336$$

y:

$$ln(Y_i/P_3) = -6.55 + 0.91ln(X_i) + 0.51ln(P_1/P_3) + 0.09ln(P_2/P_3) + \hat{\epsilon}_i$$

$$\text{con } SCE = 0.364$$

donde SCE corresponde a la suma de los errores al cuadrado. En base a esta información, ¿cómo comprobaría si la restricción $\alpha_1 + \alpha_2 + \alpha_3 = 1$ es válida? Muestre sus cálculos.

40. Los profesores M.Vargas y R.Mayer han estado sumidos en una pronfunda discusión respecto de quien es le mejor profesor de econometría. Ambos dictan la misma asignatura en una prestigiosa Universidad durante el mismo semestre. Como una manera de zanjar la disputa, el profesor M.Vargas propone estimar el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + u_i$$

donde Y_i corresponde a la nota obtenida por el alumno i, X_i es el número promedio de horas que el alumno estudió para cada una de las tres pruebas (dos solemnes y un examen), Z_i es una variable muda (*dummy*) que toma el valor uno si el alumno es hombre y cero si no, y u representa un término de error bien comportado (idd con media cero y varianza homocedástica).

a) Si se quisiera dilucidar la polémica surgida entre ambos profesores, ¿qué variable se podría incroporar en el modelo? ¿Sería necesario realizar algún test para zanjar la disputa?

b) El profesor R.Mayer señala que debiese eliminarse la variable dummy hombre, pues es irrelevante y por ende, genera sesgo en la estimación del resto de los parámetros. Comente.

c) El profesor M.Vargas dice que omitir la variable habilidad (H_i) en la estimación de este modelo no genera sesgo en los parámetros, pero sí inconsistencia. Comente.

d) ¿Qué test realizaría usted para analizar si es que el modelo está bien especificado? Sea claro y explique detalladamente cómo implementaría el test.

e) Un potencial problema de esta estimación, señala M.Vargas, es que los residuos pueden ser heteroscedásticos, lo cual podría sesgar las estimaciones. Sin embargo, R.Mayer le recuerda que de igual forma los coeficientes serán consistentes, y además, la matriz de varianzas y covarianzas vendrá dada por:

$$var(\hat{\beta}^{MCO}) = \hat{\sigma}^2(X'X)^{-1}$$

Comente lo anterior.

f) Discuta cómo implementaría el test de White para detectar la presencia de heterocedasticidad.

g) ¿Cómo podría determinar si es que las mujeres tienen un mayor retorno en términos de nota por cada hora adicional de estudio?

h) Dado que la variable X podría estar medida con error (los alumnos no señalaron de manera precisa el número promedio de horas que estudiaron para cada una de las pruebas), entonces las estimaciónes serán inconsistentes. Comente.

41. Considere la estimación de una regresión de salarios (en niveles, medidos en dólares por hora) sobre las variables dummies de raza, sexo y región geográfica, tal como se describe a continuación:

Variable dependiente: salario por hora, en US

	Regresión 1	**Regresión 2**
Constante	7,1653 (0,2222)	7,6009 (0,3524)
No-blanco	-0.6218 (0,6832)	-0.4107 (0,6799)
Mujer	-2,539 (0,3205)	-2,5648 (0,3179)
No-blanco-mujer	0,2332 (1,0021)	-0,2191 (0,9966)
Centro-norte		-0,6022 (0,4373)
Sur		-1,0013 (0,4064)
Oeste		0,4389 (0,4849)
Número de observaciones	526	526
R^2	0,1175	0,1388
R^2	0,1124	0,1288
SC residuos	6318,7635	6166,3478
Test F Global	23,18	13,94
p-value F	0,0000	0,0000
log-L	-1433,0604	-1433,0604

donde $No - blanco$ es una variable dummy que toma el valor 1 si el individuo no es de raza blanca y 0 si lo es. $Mujer$ es una variable dummy que toma el valor 1 si el individuo es mujer y 0 si no lo es. $No - blanco - mujer$ es una dummy interactiva entre $No - blanco$ y $Mujer$. $Centro - norte$ es una dummy que toma el valor 1 si el individuo vive en la región Centro o Norte del país, y 0 si no vive en estas regiones. Sur es una dummy que toma el valor 1 si el individuo vive en la región Sur del país, y 0 si no vive en esta región. $Oeste$ es una dummy que toma el valor 1 si el individuo vive en la región Oeste del país, y 0 si no vive en esta región.

Entre paréntesis aparece el error estándar de las estimaciones de MCO. Adicionalmente, se dispone de la matriz de varianzas y covarianzas de los parámetros estimados en la segunda regresión:

	No-blanco	Mujer	no-blanco-mujer	Centro-norte	Sur	Oeste
No-blanco	0,4622					
Mujer	0,0080	0,101				
No-blanco-mujer	0,1230	0,0011	0,9932			
Centro-norte	0,1622	0,0009	0,022	0,1912		
Sur	0,0053	0,0009	0,166	0,111	0,1651	
Oeste	0,0222	0,0100	0,013	0,112	0,1006	0,2351

Con respecto a la regresión 1:

a) Las dummies incorporadas (no-blanco, mujer y no-blanco-mujer), permiten dividir la muestra en cuatro grupos, ¿cuáles son estos grupos?

b) Describa cómo los respectivos coeficientes definen los valores medios de cada uno de los respectivos grupos.

c) ¿Cuál es el salario medio para las mujeres blancas? ¿Existe suficiente información para calcular el error estándar de esta media?

Con respecto a la regresión 2:

a) ¿Cómo cambia la interpreteación de las dummies incorporadas (no blanco, mujer y no-blanco mujer), cuando se agregan las variables dummies regionales?

b) ¿Cuál es el diferencial salarial asociado a vivir en el Sur versus vivir en el Oeste? Encuentre el error estándar de esta diferencia.

c) Compute el test F que permite testear si las variables dummies regionales son, en conjunto significativas.

42. En su papel de gerente de Recursos Humanos de la prestigiosa compañía de seguros SALVADO el año pasado usted decidio financiar un curso de capacitación, el cual permitiera potenciar las habilidades de su fuerza de venta, es decir, de los ejecutivos dedicados a vender los distintos productos que ofrece la

compañía. Dado que se trataba de un programa piloto, y que los recursos eran limitados, usted decidió sortear de manera aleatoria los cupos disponibles. Es así como solo el 20 de los ejecutivos de venta pudieron recibir este programa de capacitación, y usted sabe en concreto con nombre y apellido quienes participaron de éste. Con el objetivo de evaluar la efectividad que tuvo dicho programa ha decidido recopilar la siguiente información de sus vendedores para el año 2009:

- Ventas anuales (en unidades de fomento)
- Género (sexo) del vendedor
- Edad (años)
- Experiencia como ejecutivo de venta (años)
- Si tiene o no educación superior completa (universitaria, centro de formación técnica o instituto profesional).

a) Escriba el modelo econométrico que estimaría en base a la información disponible para evaluar la efectividad que tuvo el programa de capacitación. Defina cuidadosamente cada uyna de las variables del modelo.

b) Explique detalladamente cómo implementaría el test de White para detectar presencia de heteroscedasticidad.

c) ¿Qué consecuencia tendría el hecho de que sus errores fueran heteroscedásticos?

d) Si la simpatía es un factor fundamental al momento de concretar una venta, ¿cómo influye el hecho de que usted la haya omitido en el análisis econométrico?

e) Explique cómo podría testar si es que el modelo está correctamente especificado.

f) ¿Cómo modificaría el modelo planteado en (a) para incorporar el hecho de que las mujeres tienen un mayor premio asociado por cada año adicional de experiencia?

g) ¿Cómo modificaría el modelo planteado en (a) para incorporar el hecho de que las hombres tienen un mayor premio asociado al hecho tener educación superior completa?

43. Supona que se desea modelar el estado de salud de un individuo, para lo cual el econometrista plantea el siguiente modelo teórico:

$$Salud_i = b_0 + b_1 edad_i + b_2 peso_i + b_3 altura_i + b_4 mujer_i + b_5 trabajo_i +$$

$$b_6 ejercicio_i + b_7 emocional_i + u_i$$

donde *Salud* es un indicador de calidad de vida relacionado con la salud, *trabajo* corrresponde a las horas trabajadas semanalmente, *ejercicio* son las horas de ejercicio semanales y *emocional* corresponde al grado de bienestar emocional del individuo. Por otro lado, asuma lo siguiente:

$$X_1 = [i\ edad\ peso\ altura\ mujer\ trabajo\ ejercicio]$$

$$\beta_1 = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \end{bmatrix}$$

$$X_2 = [emocional]$$

$$\beta_2 = [b_7]$$

donde i corresponde a un vector columna de unos (1). En base a esta información se pide lo siguiente:

a) Asuma que el analista no posee en sus datos la varible *emocional* y por lo tanto deberá omitirla de su estimación final. Pruebe que el estimador MCO para β_1 está sesgado. ¿De qué depende este sesgo?

b) Suponga que $X = (X_1\ X_2)$, y $\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$ (en esta ocasión se tiene la variable *emocional* en la base de datos). El investigador desea incluir una variable irrelevente $X_3 = (distcasa)$, que corresponde a la distancia (en kms.) desde la casa al gimnasio más cercano. Probar que el estimador MCO para β es insesgado.

c) Con la información que se entregó en la parte (a), pruebe que si X_2 es independiente de X_1, entonces β_1 es insesgado.

44. Suponga que usted cuenta con información sobre la felicidad de las personas (B), la cual es una variable dummy (muda) que puede tomar solo dos valores: 0 si el individuo se clasifica como "no feliz" y 1 si el individuo dice "ser feliz". Por otro lado, asuma que tiene información para estos mismos individuos sobre un conjunto de variables que afectan la felicidad de las personas. En concreto, usted tiene información sobre la edad de los individuos, género, años de escolaridad, ingreso familiar, y estado de salud. En base a esta información se puede plantear el siguiente modelo econométrico:

$$B_i = \alpha_1 + \alpha_2 E_i + \alpha_3 M_i + \alpha_4 S_i + \alpha_5 I_i + \alpha_6 H_i + \epsilon_i$$

donde E es la edad de las personas, M es una variable dummy que toma el valor 1 si el individuo es mujer y cero si no, S son los años de escolaridad, I es el ingreso familiar (medidos en miles de pesos al año), H es una variable dummy que toma el valor 1 si la persona goza de buena salud y cero si no, y ϵ es un shock estocástico que captura todos aquellos factores que afectan la felicidad de las personas pero que no son observables por el econometrista.

a) ¿Es posible estimar el modelo econométrico planteado por mínimos cuadrados ordinarios (MCO)? Justifique detalladamente su respuesta.

b) En el contexto de MCO, ¿cuál es la probabilidad de que un individuo sea feliz?

c) ¿Cuál es el efecto marginal que tiene el ingreso familiar sobre la probabilidad de ser feliz?

d) Mencione y explique dos críticas a la estimación por MCO del modelo planteado. En vista de esto, ¿qué método de estimación es más apropiado?

Suponga ahora que existe una variable latente B^* (que no es observable) que representa el nivel de felicidad de las personas. Luego el modelo ahora es el siguiente:

$$B_i^* = \gamma_1 + \gamma_2 E_i + \gamma_3 M_i + \gamma_4 S_i + \gamma_5 I_i + \gamma_6 H_i + \zeta_i$$

donde $\zeta \sim N(0,1)$. Lo que el econometrista observa es la variable dictotómica B que como se dijo anteriormente toma el valor 0 si el individuo se reporta como no feliz y 1 si el individuo dice ser feliz. En concreto el econometrista observa lo siguiente:

$$B_i = \begin{cases} 1 \ si \ \mathrm{B}_i^* \ > 0 \\ 0 \ si \ \mathrm{B}_i^* \leq 0 \end{cases}$$

e) A partir de la nueva ecuación, ¿cuál es la probabilidad de que el individuo sea feliz? ¿Cuál es la probabilidad de que el individuo sea no feliz? Ayuda: La función de densidad de una distribución normal estandar es:

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Por otro lado, la función de distribución acumulada es:

$$\phi(q) = \int_{-\infty}^{q} \phi(t)dt$$

f) ¿Cuál es el efecto marginal que tiene ahora el ingreso familiar sobre la probabilidad de ser feliz y cómo se compara con el encontrado en la parte (c)?

g) Encuentre la función de log-likelihood de este problema.

h) Explique cómo encontraría la matriz de varianzas y covarianzas de los coeficientes estimados.

45. Considere el siguiente modelo de regresión lineal con k variables:

$$Y = X\beta + u$$

donde Y es un vector columna de dimensión nx1, X es una matriz de dimensión nxk y u es un vector columna de dimensión nx1. Asuma que el shock estocástico u presenta las siguientes propiedades:

$$E(u_i) = \sigma_i^2 \ para \ i = 1, 2, \ldots, n$$

$$E(u_i u_j) = 0 \ \forall \ i \neq j$$

En base a esta información es posible encontrar la matriz de varianzas y covarianzas del término de error:

$$E(uu') = \sigma^2 \Omega$$

a) Derive formalmente el estimador de mínimos cuadrados ordinarios (MCO) de β. ¿Es insesgado el estimador de MCO? Demuestre.

b) Derive la matriz de varianzas y covarianzas del estimador de MCO de β. ¿Es MELI el estimador de MCO? Explique.

c) ¿Es posible transformar este modelo de perturbaciones no esféricas a uno con perturbaciones esféricas? Explique lo más detalladamente el procedimiento que seguiría. Ayuda: Recuerde la descomposición espectral de la matriz Ω.

d) Encuentre la matriz de varianzas y covarianzas del estimador de mínimos cuadrados generalizados (MCG).

e) ¿Por qué se dice que el estimador MCO es un caso particular del estimador MCG?

f) ¿Cómo podría estimar el modelo si la matriz Ω fuera desconocida? Explique.

46. Considere las siguientes especificaciones econométricas:

Modelo 1:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \epsilon_i$$

Modelo 2:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \epsilon_i$$

a) ¿Es $\hat{\beta}_2$ insesgado? Demuestre.

b) ¿Es $\hat{\beta}_2$ eficiente? Explique claramente su respuesta.

Suponga ahora que el modelo verdadero es el Modelo 2, pero se estima el Modelo 1 por MCO. Bajo estas circunstancias:

c) ¿Son los estimadores de las pendientes insesgados? Demuestre.

d) ¿Son los estimadores de las pendientes eficientes? Explique claramente su respuesta.

47. El poblado sureño de Meaogo esta situado a orillas del río 'Crecedorcito'. El año 2009 hubo dos serias inundaciones en Meaogo por efecto de las crecidas del río. Suponga que el número de inundaciones por año en esta localidad (x) sigue una distrubición Poisson con parámetro λ desconocido:

$$p(x;\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} \; para \; x = 0, 1, 2, \ldots \\ 0 \; para \; cualquier \; otro \; valor \end{cases}$$

a) Si tiene una sola observación (la del año 2009), ¿puede estimar el valor desconocio de ? Explique.

b) Si su respuesta a la pregunta anterior es afirmativa, escriba la función de verosimilitud. Si es no, explique qué información adicional requeriría para encontrar analíticamente esta función.

Suponga ahora que λ sólo puede tomar uno de los siguientes valores: 1, 1.5, 2, 2.5, 3.

c) ¿Puede determinar con cuál de estos valores usted se quedaría si le preguntan su mejor estimación del parámetro poblacíónal λ?

48. Suponga que usted se desempeña como gerente de Recursos Humanos de una importante empresa multinacional, y el directorio le ha solicitado la elaboración de un estudio en donde analice los determinantes de los salarios de los trabajadores de la compañía. Usted cuenta con información relativa a los salarios mensuales (W) de todos los empleados, además de algunas otras características de los inviduos como edad (E), antigüedad en la empresa (A) y escolaridad (S). Estas dos últimas variables se encuentran medidas en número de años. En base a esta información, explique claramente cómo usted abordaría los siguientes requerimientos:

a) Especifique un modelo para la determinación de los salarios de los trabajadores de la empresa. Defina claramente cada una de las variables involucradas. Realice los supuestos que estime pertinentes para el término de error (u).

b) El directorio está pensando en contratar un profesional universitario de 30 años de edad, ¿Qué salario se le debería ofrecer para ser consistenten con la estructura de remuneraciones de la empresa?

c) Si usted contara con información relativa al sexo (género) de los trabajadores, ¿De qué manera la inocorporaría en la estimación? Sea claro.

d) ¿Cómo incorporaría el hecho de que los hombres reciben un mayor premio salarial que las mujeres por cada año adicional de antigüedad en la empresa?

e) Si usted cree que el impacto que tiene cada año adicional de escolaridad sobre los salarios de las personas no es lineal (es cóncavo), ¿cómo incorporaría usted este hecho a su modelo?

f) La habilidad (H) es un determinante fundamental de los salarios de las personas, pero como usted no cuenta con información acerca de la habilidad de ellos ha debido omitirla. ¿Qué implicancias tendrá esto sobre los parámetros estimados?

g) Suponga que usted tiene sospecha de la existencia de heteroscedasticidad en sus residuos, ¿Cómo implementaría el test de White para testear la hipótesis nula de errores homoscedásticos?

h) Si usted se encuentra que los errores son efectivamente heteroscedásticos, ¿Qué consecuencias tendría esto para la estimación de mínimos cuadrados ordinarios (MCO)?

i) Si usted conociera la forma que tiene la heterocedasticidad, ¿qué método de estimación implementaría? Muestre claramente la fórmula de dicho estimador.

j) Si tiene sospecha de un elevado grado de correlación entre las variables años de escolaridad y años de experiencia en la empresa (antigüedad), ¿Qué implicancias tendría solo esto para su estimación?

49. Dado el siguiente modelo:

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

$$u_t = \rho u_{t-1} + e_t$$

donde $e_t \sim N(0, \sigma^2)$. Además, dispone de las siguientes observaciones:

t	1	2	3	4	5	6	7	8
X	4	6	10	12	14	16	20	22
Y	22	26	32	34	40	46	46	50

Obtenga una estimación eficiente de los parámetros de β_0 y β_1, sabiendo que $\rho = 0,5$.

50. Considere el siguiente modelo:

$$y_i = x_i'\beta + u_i$$

con $i = 1, 2, \ldots, n$; donde x_i es un vector de dimensión 1xk, y β es un vector de dimensión kx1. Por otro lado, u, es un término de error que se distribuye normal con media cero y varianza σ^2. Por lo tanto:

$$f(u_i) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{u_i^2}{2\sigma^2}\right)$$

a) Encuentre la función $l = ln(L)$, donde L es la verosimilitud de la muestra.

b) Demuestre formalmente que el estimador de máxima verosimilitud (MV) de β viene dado por:

$$\hat{\beta}_{MV} = (X'X)^1 X'y$$

c) ¿Bajo qué condiciones el estimador de MV coincide con el estimador de mínimos cuadrados ordinarios (MCO)?

d) Demuestre formalmente que el estimador de MV para σ^2 viene dado por:

$$\hat{\sigma}^2_{MV} = \frac{\hat{u}'\hat{u}}{n}$$

e) Encuentre la matriz de varianzas y covarianzas de los estimadores de MV de β y σ^2.

f) Suponga ahora que se tiene el siguiente modelo alternativo:

$$y_i = z_i'\beta + \epsilon_i$$

donde ϵ se distribuye normal con media cero y varianza homocedástica. Por otro lado, z es un vector que contiene solo algunas variables de las contenidas en x. ¿Cuál modelo seleccionaría usted? Explique.

51. Usando una muestra de datos para Chile se ha estimado la siguiente ecuación por mínimos cuadrados ordinarios (MCO):

$$\widehat{psu}_i = \underbrace{1028,10}_{6,29} + \underbrace{19,30}_{3,83} clase_i - \underbrace{2,19}_{0,53} clase_i^2 - \underbrace{45,09}_{4,29} mujer_i -$$

$$\underbrace{169,81}_{12,71} indigena_i + \underbrace{62,31}_{18,15} (mujer_i \cdot indigena_i)$$

con $n = 4131$ y $R^2 = 0.0858$. Debajo del coeficiente se reporta el desvío estándar del estimador MCO. La variable *psu* mide la puntuación de la prueba de selección universitaria del alumno, *clase* es el tamaño de la generación a la cual pertenece el alumno (medido en cientos de alumnos), *mujer* es una variable *dummy* (muda) que toma el valor de 1 si el alumno es de sexo femenino y 0 si es de sexo masculino, e *indigena* es una variable *dummy* que toma el valor 1 si el alumno pertenece a una etnia indígena y 0 si no.

a) ¿Hay evidencia fuerte de que la variable $clase^2$ debe incluirse en el modelo? De acuerdo con esta ecuación, ¿cuál sería el tamaño óptimo de la clase?

b) Manteniendo fijo el tamaño de clase, ¿cuál es la diferencia de puntajes estimada en el puntaje PSU entre las mujeres no indígenas y los hombres no indígenas? ¿Es estadísticamente significativo esta diferencia?

c) ¿Cuál es la diferencia estimada de puntajes PSU entre hombres indígenas y hombres que no son indígenas? Contrastar la hipótesis nula de que no hay diferencia en sus puntajes, contra alternativa de que sí la hay.

d) ¿Cuál es la diferencia estimada de puntajes PSU entre mujeres indígenas y mujeres que no lo son indígenas? ¿Qué se necesitaría para hacer contrastar la hipótesis de que la diferencia es estadísticamente significativa?

52. Los profesores Nicholas Bloom y John Van Reenen desean ajustar un modelo que explique la calidad de gestión en las empresas (z_i) en función de la edad de la firma (*edad*), si el gerente general de la compañía posee un título MBA (MBA), un índice que mide el grado de competencia de mrecado (l), una variable dicotómica que toma el valor 1 si la empresa es pública (*public*) y 0 en otro caso, y un shock estocástico u con media 0. De esta manera, estimano por mínimos cuadrados ordinaios (MCO) la siguiente especificación para una muestra de 726 empresas:

$$z_i = \alpha + \beta edad_i + \gamma MBA_i + \delta l_i + \tau public_i + u_i$$

Los resultados se muestran en la siguiente tabla:

Variable dependiente: calidad de gestión (puntaje)
Estimación por MCO

(1) Parámetro	(2) Coeficiente	(3) S.D	(4) S.D Robusta
	-0,001	0,001	0,001
	0,030	0,008	0,007
	0,867	0,497	0,482
	0,103	0,052	0,053
	-0,863	0,479	0,463
White-test($nR^2 \sim \chi^2_{13}$)	27,710	(P-value = 0,01)	

La columna (3) muestra las desviaciones estándar de las estimaciones por MCO usando la fórmula habitual de la varianza $var(\hat{\beta}|X) = \sigma^2(X'X)^{-1}$, mientras que la columna (4) registra las correspondientes desviaciones estándar utilizando la matriz de varianzas y covarianzas de White.

a) Refiérase a la evidencia que muestra el test de White. Explique cómo se calcula, su hipótesis nula y las conclusiones a la luz de los resultados mostrados en el enunciado.

b) Explique cómo se calcula la column (4). ¿Cuál es la utilidad de la correción por la matriz de White?

c) Con la evidencia del enunciado, ¿es posible afirmar que sólo por el hecho de que una empresa sea pública, entonces, ello debiera tener algún impacto en la calidad de la gestión de ésta? Argumente de manera clara su respuesta, indicando los datos que usó para resolver esta interrogante.

d) Explique el procedimiento para obtener el estimador de Mínimos Cuadrados Generalizados (MCG)y obtenga una expresión para $\hat{\beta}_{MCG}$.

e) Demuestre que MCG es un estimador eficiente.

53. Considere el siguiente modelo que resume la verdadera relación que existe entre la variable dependiente (Y) y un conjunto de variables independientes (X_1):

$$Y = X_1\beta_1 + u$$

donde Y es un vector de dimensión nx1, X_1 es una matriz de dimensión nxk, β_1 es un vector de parámetros de dimensión k_1x1 y u es un vector de dimensión $nx1$ de shocks estocásticos. Asuma que dichos shocks se distribuyen normal con media cero, y matriz de varianzas y covarianzas dada por $E(uu') = \sigma^2 I_n$.

Sin embargo por alguna razón el econometrista decidió incorporar un conjunto de variables irrelevantes (k_2) en la estimación del modelo. En concreto, estimó lo siguiente:

$$Y = X_1\beta_1 + X_2\beta_2 + u$$

donde X_2 es una matriz de dimensión nxk_2 y β_2 es un vector de parámetros de dimensión k_2x1. La estimación por mínimos cuadrados ordinarios de este modelo viene dada por:

$$\hat{\beta}_j = (X_j' M_i X_j)^{-1} X_j' M_i Y$$

con $j \neq i = 1, 2$

$$M_i = I_n - X_i(X_i' X_i)^{-1} X_i'$$

con $i = 1, 2$

a) Demuestre que el estimador de mínimos cuadrados ordinarios (MCO) para β_1 es insesgado y consistente.

b) Encuentre la matiz de varianzas y covarianzas del estimador MCO de β_1.

c) Demuestre que la inclusión de variables irrelevantes genera una pérdida de eficiencia al estimar β_1.

54. El sicólogo de una prestigiosa Universidad está interesado en analizar los determinantes del desempeño de los estudiantes durante su primer año de educación superior. En particular, desea evaluar si un novedoso programa de talleres de preparación para la educación superior (TPES), impartido por la misma Universidad a alumnos de cuarto medio después de la jornada escolar, afecta o no el rendimiento durante el primer año de carrera. Para ello, cuenta con una muestra de 50 alumnos (de un total de 200 alumnos admitidos) aceptados en la carrera de Administración de Negocios para el año 2009 y que asistieron a TPES. La base de datos a la que tiene acceso contiene la siguiente información:

    Ranking: Lugar del almuno, en percentiles, respecto del total de alumnos de su generación de acuerdo a su promedio de notas durante el primer año de Universidad. 100 corresponde al mejor alumnos, 1 corresponde al peor.

    Mate: Resultado de la prueba de diagnóstico de matemáticas administrada a los alumnos durante la primera semana de clases en la Universidad.

    Horas: Horas de asistencia semanal promedio de los alumnso durante cuarto medio a los talleres de preparación para la educación superior (TPES).

    Estadística descriptiva para estos datos se presenta en la siguiente tabla:

Ranking		*Mate*		*Horas*	
Media	0,53	Media	4.89	Media	30.10
Error típico	0.03	Error típico	0.38	Error típico	0.29
Mediana	0.49	Mediana	4.47	Mediana	30.19
Desv. estándar	0.22	Desv. estándar	2.70	Desv. estándar	2.06
Var. de la muestra	0.05	Var. de la muestra	7.27	Var. de la muestra	4.23
Curtosis	-1.08	Curtosis	-1.31	Curtosis	-0.15
Coef. de asimetría	0.37	Coef. de asimetría	0.24	Coef. de asimetría	0.33
Mínimo	0.21	Mínimo	1.07	Mínimo	26.22
Máximo	1	Máximo	9.94	Máximo	35.02
Cuenta	50	Cuenta	50	Cuenta	50

    El sicólogo piensa que debe controlar por el resultado de la prueba de diagnóstico de matemáticas puesto que, según otros estudios, los alumnos con más habilidades matemáticas tienden a tener mejores resultados en la Universidad. De acuerdo a esta hipótesis, estima la siguiente regresión:

$$Ranking_i = \beta_1 + \beta_2 Mate_i + \beta_3 Horas_i + \upsilon_i$$

Los resultados de la estimación se presentan en la siguiente tabla 2:

a) De acuerdo a estos resultados, ¿qué opina usted de la efectividad del programa TPES?.

Tabla 2: Regresión de Ranking en Mate y Horas

Estadísticas de la regresión	
Coeficiente de correlación multiple	0.936
Coeficiente de determinación	0.876
R ajustado	0.871
Error típico	0.079
Observaciones	50

	Coeficientes	*Error típico*	*Estadistico t*	*valor p*
Intercepto	-0.0392	0.1657	0.2367	0.8139
Mate	0.0750	0.0043	17.6003	0.0000
Horas	0.0067	0.0056	1.1941	0.2384

El sicólogo no queda satisfecho con estos resultados y decide estimar en forma separada los efectos de Mate y Horas sobre el rangkin que obtienen los alumnos durante su primer año de Universidad. Los resultados son los siguientes:

Tabla 3: Regresión de Ranking en Mate

Estadísticas de la regresión	
Coeficiente de correlación multiple	0.934
Coeficiente de determinación	0.872
R ajustado	0.870
Error típico	0.079
Observaciones	50

	Coeficientes	*Error típico*	*Estadistico t*	*valor p*
Intercepto	0.1567	0.0234	6.7043	0.0000
Mate	0.0761	0.0042	18.1199	0.0000

Tabla 4: Regresión de Ranking en Horas

Estadísticas de la regresión	
Coeficiente de correlación multiple	0.246
Coeficiente de determinación	0.060
R ajustado	0.041
Error típico	0.215
Observaciones	50

	Coeficientes	*Error típico*	*Estadistico t*	*valor p*
Intercepto	-0.2605	0.4504	-0.5783	0.5658
Horas	0.0262	0.0149	1.7552	0.0856

El sicólogo queda muy confundido, puesto que de acuerdo a los resultados presentandos en las tablas 3 y 4, ambas variables, en forma individual, afectan significativamente, al menos al 10 por ciento de significancia, el desempeño de los estudiantes durante su primer año de Universidad. Sin saber que hacer, decide graficar Mate contra Horas, pero el gráfico no le ayuda mucho a resolver sus dudas:

b) ¿Qué cree Usted que puede estar ocurriendo? Explique claramente su respuesta.

c) Si se observa los coeficientes estimados para la variables $Horas$ en la tabla 2 (0.0067) y en la tabla 4 (0.0262), se dará cuenta que el coeficiente de esta variable cuando no se controla por Mate es de un orden de magnitud superior a que cuando sí se control. ¿Puede usted explicar por qué se produce un cambio tan grande en el coeficiente estimado de Horas cuando se pasa de un modelo de regresión múltiple, donde se controla por $Mate$ a un modelo simple, donde no?

d) De acuerdo a la evidencia presentada, ¿piensa usted que $Horas$ pertence al modelo? ¿Puede testarlo formalmente? Si puede, hágalo. El test que utilizaría, ¿sigue siendo valido en el contexto de la pregunta? Explique. Si no puede testear formalmente si $Horas$ pertenece al mdoelo, explique qué información adicional requeriría para sí poder hacerlo.

55. Suponga que usted está interesado en estimar la función de demanda de la bencina de 95 octanos en la región metropolitana. Por ello usted plantea el siguiente modelo econométrico:

$$ln(Q_i) = \alpha_1 + \alpha_2 ln(P_i) + +u_i$$

donde Q es la cantidad de bencina de 95 octanos vendida en la comuna i, P es el precio promedio de venta de la bencina (entre los distintos distribuidores) en la comuna i, y u es una perturbación esférica con $E(u_i) = 0$ para todo $i = 1, 2, \ldots, n$.

a) ¿Que representa α_2? ¿Qué signo esperaría para este parámetro?

b) Es esperable que aquellas comunas de mayor nivel socioeconómico consuman una mayor cantidad de bencina. Es por ello que usted desea incorporar en el modelo el nivel de ingreso familiar promedio de la comuna (Y_i) en la especificación econométrica. En particular, usted plantea ahora el siguiente modelo econométrico:

$$ln(Q_i) = \alpha_1 + \alpha_2 ln(P_i) + \alpha_3 Y_i + u_i$$

En vista de esto, ¿qué ocurriría con la estimación por minímos cuadrados ordinarios (MC0) del primer modelo considerando que el segundo modelo sea el verdadero? En particular refiérase al estimador de MCO para α_2.

c) Explique cómo implementaría el test de White para testear la presencia de errores heteroscedásticos en el segundo modelo. No olvide plantear la hipótesis nula, la distribución dle test, y los criterios de rechazo.

d) Considere la situación en que usted encuentra evidencia estadística de presencia de errores heterocedásticos en el segundo modelo. ¿Qué consecuencias tiene esto para la estimación por MCO? Refiérase a cómo estimaría los parámetros y la matriz de varianzas y covarianzas.

e) Suponga que usted conoce exactamente la forma que toma la heteroscedasticidad en el modelo. En particular usted sabe que:

$$var(u_i) = E(u_i^2) = \gamma Y_i^2$$

con

$$\gamma > 0$$

Si usted conociera el valor exacto de γ, ¿cómo estimaría los parámetros del modelo? Explique.

f) Asuma ahora que usted no conoce la forma que toma la heterocedasticidad, solo sabe que está presente. ¿Cómo estimaría los parámetros del segundo modelo y la matriz de varianzas y covarianzas?

56. La variable *smokes* es una variable dummy que toma el valor uno (1) si la persona fuma y cero si no. Utilizando una base de datos se estimó el siguiente modelo de probabilidad lineal para identificar los determinantes de la probabilidad de fumar de los individuos (mínimos cuadrados ordinarios):

Determinantes de la probabilidad de fumar

(Variable dependiente: *smokes***)**

Variable	**Coeficiente**	**Error estándar (sin tomar en cuenta la heterocedasticidad)**	**Error estándar robusto a la heterocedasticidad**
Constante	0.656	0.855	0.856
ln (*cigprig*)	-0.069	0.204	0.207
ln (*income*)	0.012	0.026	0.026
educ	-0.029	0.006	0.005
age	0.020	0.006	0.005
age^2	0.00026	0.00006	0.00006
restaurn	- 0.101	0.039	0.038
white	-0.026	0.052	0.050
N	807		
R^2	0.062		

La variable *white* toma el valor uno (1) si el encuestado es blanco y cero en caso contrario; $ln(cigprig)$ es el logaritmo del precio del paquete de cigarrillos, ln(*income*) es el logaritmo del ingreso anual de la persona, *educ* son los años de escolaridad, *age* y age^2 son la edad y su cuadrado, y *restaurn* es una variable dummy que toma el valor uno (1) si la persona reside en una zona geográfica con restricciones al consumo del tabaco en los restaurantes.

a) ¿Hay diferencias importantes entre las dos expresiones de error estándar? Explique cómo se calcula cada una de ellas.

b) Manteniendo los demás factores constantes, si la educación se incremente en cuatro años, ¿qué le ocurre a la probabilidad estimada de fumar?

c) ¿A partir de qué edad el tener un año más reduce la proabilidad de fumar?

d) Interprete el coeficiente de la variable $restaurn$.

e) Una persona de la muestra tiene las siguientes características: $cigpric = 67.44$; $income = 6,500$; $educ = 16$; $age = 77$; $restaurn = 0$; $white = 0$ y $smokes = 0$. ¿Cuál es la probabilidad estimada de fumar de esta persona?

57. Considere la siguiente función de consumo agregado:

$$CONS_t = \beta_0 + \beta_1 Y_t^{\beta_2} + \beta_3 CONS_{t-1} + \epsilon_t$$

donde $CONS$ es el consumo agregado e Y corresponde al ingreso agregado. Considere los siguientes resultados de la estimación de este modelo:

Dependent Variable: CONS
Sample: 1963 1982
Included observations: 20

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	-24.08890	12.39205	-1.943900	0.0677
PIB	0.644764	0.010197	63.22833	0.0000
R-squared	0.995518	Mean dependent var		747.8295
Adjusted R-squared	0.995269	S.D dependent var		138.1837
S.E of regression	9.504876	Akaike info criterion		7.436126
Sum squared resid	11790.25	Schwarz criterion		9.456055
Log likelihood	-72.36126	F-statistic		3997.822
Durbin-Watson stat	1.123553	Prob(F-statistic)		0.000000

Dependent Variable: CONS
Sample: 1963 1982
Included observations: 19 after adjustments

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	-7.695750	11.44429	-0.672453	0.5109
PIB	0.400147	0.062718	6.380139	0.0000
CONS (-1)	0.380727	0.094786	4.016699	0.0010
R-squared	0.997583	Mean dependent var		759.6800
Adjusted R-squared	0.997281	S.D dependent var		131.1138
S.E of regression	6.836788	Akaike info criterion		6.826452
Sum squared resid	747.8667	Schwarz criterion		6.975574
Log likelihood	-61.85130	F-statistic		3302.055
Durbin-Watson stat	1.356622	Prob(F-statistic)		0.000000

Dependent Variable: CONS
Sample: 1963 1982
Included observations: 19 after adjustments
CONS = C(1) + C(2)*PIB$^C(3) + C(4) * CONS(-1)$

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C(1)	75.08795	92.24997	0.813962	0.4284
C(2)	0.079656	0.168688	0.472211	0.6436
C(3)	1.206099	0.268127	4.498233	0.0004
C(4)	0.358588	0.096585	3.712687	0.0021
R-squared	0.997762	Mean dependent var		759.6800
Adjusted R-squared	0.997315	S.D dependent var		131.1138
S.E of regression	6.793984	Akaike info criterion		6.854616
Sum squared resid	692.3732	Schwarz criterion		7.053445
Log likelihood	-61.11885	Durbin-Watson stat		1.501228

Dependent Variable: RESID

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	3.586437	11.32972	0.316551	0.7563
PIB	-0.051439	0.072151	-0.712929	0.4876
CONS (-1)	0.080272	0.109540	0732811	0.4758
RESID (-1)	0.541506	0.325190	1.665199	0.1181
RESID (-2)	-0.487386	0.273797	-1.780101	0.0968
R-squared	0.248323	Mean dependent var		-1.11E-13
Adjusted R-squared	0.033559	S.D dependent var		6.445785
S.E of regression	6.336707	Akaike info criterion		6.751530
Sum squared resid	562.1540	Schwarz criterion		7.000066
Log likelihood	-59.13953	F-statistic		1.156257
Durbin-Watson stat	2.032955	Prob(F-statistic)		0.371273

Dependent Variable: RESID^2

Variable	Coefficient	Std.Error	t-Statistic	Prob.
C	-289.1809	622.6691	-0.464421	0.6495
PIB	1.323682	4.306069	0.307399	0.7631
PIB^2	- 0.000594	0.001627	-0.364959	0.7206
CONS(-1)	-1.292785	6.114797	-0.211419	0.8356
CONS(-1)^2	0.001022	0.003733	0.273821	0.7882
R-squared	0.040698	Mean dependent var		39.36140
Adjusted R-squared	-0.233389	S.D dependent var		47.57952
S.E of regression	52.84088	Akaike info criterion		10.99338
Sum squared resid	39090.22	Schwarz criterion		11.24192
Log likelihood	-99.43712	F-statistic		0.148485
Durbin-Watson stat	2.029878	Prob(F-statistic)		0.960662

En base a esta información, responda las siguientes preguntas:

a) Pruebe la hipótesis de que la función de consumo agregada es lineal, al 95 por ciento de confianza.

b) Pruebe la hipótesis de que la propensión marginal a consumir de largo plazo ($PMCLP$) es constante e igual a la propensión marginal a consumir de corto plazo (PMC), al 95 por ciento de confianza. Usted sabe que:

$$PMC = \beta_1 \beta_2 Y^{\beta_2 - 1}$$
$$PMCLP = \beta_1 \beta_2 Y^{\beta_2 - 1} / (1 - \beta_3)$$

c) Suponga que la función de consumo es lineal. Realice tests de autocorrelación y de heterocedasticidad sobre los residuos de la regresión correspondiente. En base a dichos tests estadíticos, ¿las desviaciones estándar reportadas en la regresión son las apropiadas o cuáles deberían utilizarse en su lugar? Justifique su respuesta teóricamente.

Referencias

[1] Greene, W. (1993) *Econometric Analysis*. Prentice Hall, tercera edición.

[2] Gujarati, D. (2004) *Econometría*. McGraw Hill, cuarta edición.

[3] Heij, C., P. de Boer, P. Hans, T. Kloek y H. K. van Dijk. (2004) *Econometric methods with applications in business and economics*. Oxford University Press.

[4] Hendry, D. y M Morgan. (1995) *The foundations of econometric analysis*. Cambridge University Press.

[5] Johnston, J. y J. DiNardo. (1997) *Econometric Methods*. McGraw Hill, cuarta edición.

[6] Kennedy, P. (2003) *A guide to econometrics*. The MIT Press Cambridge Massachusetts, quinta edición.

[7] Newey, W. y K. West (1987) *A simple positive semi-definite, heterocedasticity and autocorrelation consistent covariance matrix*. Econometrica, 55, pp.

[8] Pesarán, P. (1987) *Econometrics*. En The New Palgrave: A Dictionary of Economics, 2, London.

[9] Simon, J. (1994) *The art of forecasting: A wager.* Cato Journal 14, 159-61.

[10] White, H. (1980) *A heterocedasticity-consistent covariance matrix estimator and a direct test for heterocedasticity*. Econometrica, 48, pp. 817-838.

[11] Wooldridge, J. (2006) *Introducción a la econometría. Un enfoque moderno*. Thomson, segunda edición.