

15.1. INTRODUCCIÓN	584
15.2. ESPECIFICACIÓN DEL MODELO	589
SUPUESTOS DEL MODELO LINEAL GENERAL.....	591
15.3. ESTIMACIÓN POR MÍNIMOS CUADRADOS ORDINARIOS (MCO).....	595
15.4 PROPIEDADES DE LOS ESTIMADORES MÍNIMO CUADRÁTICOS ORDINARIOS	609
ESTIMADOR LINEAL.....	610
ESTIMADOR INSESGADO	610
ESTIMADOR ÓPTIMO.....	611
VARIANZA DEL ESTIMADOR.....	616
15.5. NORMALIDAD DE LA PERTURBACIÓN ALEATORIA	622
CASOS ESPECIALES.....	629
15.6. CRITERIO DE MÁXIMA VEROSIMILITUD	634
CONDICIONES DE SEGUNDO ORDEN.....	637
ANÁLISIS DEL SESGO	639
EL VALOR DE LA FUNCIÓN DE VEROSIMILITUD	641
CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS	643
CASO 15.1: CORRELACIÓN DE MUESTRAS	643
CASO 15.2: ESTIMACIÓN DE PARÁMETROS	643
CASO 15.3: CONSUMO DE CERVEZA Y MORTALIDAD INFANTIL.....	644
CASO 15.4: PRIMEROS PASOS EN EViews.....	645
<i>Actividades Propuestas</i>	653
BIBLIOGRAFÍA.....	654

Capítulo 15. EL MODELO LINEAL

GENERAL

En este capítulo se continúa con el proceso de investigación econométrica. En el desarrollo del mismo, el investigador ya ha aprehendido todas las etapas; tiene su problema a investigar, ha planteado y completado su tabla de datos –luego de definir sus fuentes de información y de recolectar los datos– y ha aplicado métodos básicos de análisis de la información, tanto sobre las unidades de observación como sobre las variables involucradas en su estudio. Es en este momento en el que debe especificar y estimar un modelo econométrico, para establecer relaciones entre variables y corroborar -en espacio y tiempo específico- sus teorías y modelos económicos. En este capítulo se muestran cuáles son las condiciones necesarias y los conocimientos -matemáticos y estadísticos- suficientes que el investigador debe tener para poder realizar dicha especificación.

15.1. Introducción

En el Capítulo 1 de este manual, se definió econometría como: ... *la aplicación de métodos matemáticos y estadísticos a tablas de datos, que contienen "unidades" (de observación) por características observables de las mismas ("variables"), con el propósito de dar contenido empírico a las teorías económicas planteadas en modelos, verificándolas a partir del estudio de la semejanza entre unidades y la relación entre variables, en un espacio y tiempo específico.*

Para lograr ese objetivo, se utiliza como instrumento básico un modelo -denominado *modelo lineal general*- que es una representación simplificada del mundo real. Este modelo, para ser operativo, ha de estar expresado en forma matemática.

Ejemplo 15.1 Si se desea estudiar el consumo familiar, en un período de tiempo determinado (t), la teoría económica modela el consumo en función de la renta, es decir:

$$C = f(R) \text{ con } f'_R > 0$$

Para poder trabajar con este modelo se supondrá una forma funcional para f , por ejemplo una relación lineal, que se escribirá como:

$$C_t = \alpha + \beta R_t; \quad \forall t = 1, \dots, T$$

Donde: α representa el consumo autónomo y β la propensión marginal a consumir que se supone comprendida en el intervalo $[0,1]$

En el modelo del ejemplo 15.1 se explica el consumo por medio de una variable que determine el nivel de renta.

De acuerdo a esta especificación, se consume -en un momento de tiempo- una proporción de la renta medida por βR_t , la diferencia entre ambas cifras se supone constante (α).

Este modelo de consumo se puede utilizar:

- A nivel agregado, en cuyo caso las variables C_t y R_t serán indicadores del nivel de consumo y renta agregados, para lo cual se requieren observaciones numéricas de las variables durante un periodo de tiempo t . Por lo tanto, las observaciones correspondientes a cada una de las variables es una *serie temporal*.
- A nivel desagregado, por ejemplo relacionando los gastos semanales en consumo y los ingresos de las familias, las observaciones correspondientes a cada una de las variables es un dato obtenido de una muestra de un conjunto de familias y se denominan *datos de corte transversal*.
- En forma conjunta, combinando las observaciones de una muestra de corte transversal en el tiempo, esto se denomina *datos de panel*.

Hay que tener en cuenta que, para realizar este estudio sobre el consumo o cualquier otro que requiera la especificación de un modelo lineal general, se deberá aplicar el *proceso de investigación econométrica* descrito a lo largo de los anteriores capítulos. Es decir, se deberá definir el problema a investigar -en este caso el consumo-; se planteará una tabla de datos que contendrá unidades

de observación temporales (estudio de series de tiempo), individuos (corte transversal) o una combinación de ambas (datos de panel); se diseñará la fuente de información; se recolectará la misma y, finalmente, se podrá realizar la especificación y estimación del modelo propuesto.

Ejemplo 15.2 Si se desea estimar la función de producción de una empresa, en un período de tiempo determinado (t), la teoría económica modela la producción como una función de los factores trabajo y capital:

$$P = f(L, K)$$

Donde

P , es la producción, L es el factor trabajo y K el capital.

Si la función de producción de la empresa es de la forma Cobb-Douglas:

$$P_t = AL_t^\alpha K_t^\beta; \quad \forall t = 1, \dots, T$$

El valor de la suma $\alpha + \beta$ va a determinar si la empresa tiene rendimientos a escala constantes, crecientes o decrecientes.

Al realizar un trabajo econométrico, el primer paso es formular un modelo que, aun siendo una representación simplificada de la realidad, permita reproducir los patrones de comportamiento entre las variables económicas. Generalmente, la teoría económica no suele dar muchas indicaciones de cuál es la forma funcional del modelo, por lo que han de realizarse supuestos al respecto.

El segundo paso es estimar los parámetros de interés del modelo a partir de los datos disponibles y contrastar aquellas hipótesis estadísticas y económicas que son relevantes.

Ejemplo 15.3 Si se estimó una función de producción Cobb–Douglas, puede interesar contrastar la hipótesis de que la empresa tiene rendimientos a escala constantes; es decir, que $\alpha + \beta = 1$.

Por último, el modelo econométrico estimado y validado, se puede utilizar para predecir valores futuros de las variables o tomar decisiones de política económica.

Se recurre al *análisis de regresión* para describir la relación existente entre una variable endógena o dependiente -de argumento vectorial denominada y - y un conjunto de variables exógenas, explicativas o independientes -representadas genéricamente por la matriz de regresores X -.

Estas relaciones pueden ser de carácter *deterministas* o *estocásticas*. Las primeras son las utilizadas en los modelos de los ejemplos anteriores; las segundas, como las del ejemplo 15.4, tienen en cuenta *factores aleatorios* que influyen en el comportamiento de los agentes económicos.

Ejemplo 15.4 El modelo de consumo supone que, dada una renta R_0 , todas las familias (de un conjunto de n familias) con esa renta presentan un mismo nivel de consumo, lo que es poco realista.

Para modelar este comportamiento individual se introduce un término aleatorio, la perturbación ε_i :

$$C_i = \beta_1 + \beta_2 R_i + \varepsilon_i; \quad \forall i = 1, \dots, n$$

Los parámetros del modelo son los coeficientes β_j y los que caracterizan la función de distribución de la perturbación aleatoria vectorial ϵ ; es decir, su valor medio y su matriz de varianzas y covarianzas. El conjunto de parámetros a estimar da origen a lo que se denomina un vector paramétrico: $\mathbf{w} = [\beta_j \ \sigma_\epsilon^2]; \quad \forall j = 1, \dots, k$

Así, el modelo lineal general se puede estudiar siguiendo cuatro pasos:

- Especificación del modelo econométrico.
- Estimación de los parámetros.
- Validación del modelo.
- Predicción de valores futuros.

En este capítulo se estudiarán los dos primeros, dejando para el próximo los restantes. Cada paso se ilustrará con una aplicación numérica utilizando los datos de la tabla de datos 15.1, los que fueron obtenidos de una muestra para las variables Y , X_2 y X_3 en 5 observaciones de tiempo.

Tabla 15.1

Observaciones	Y_t	X_{2t}	X_{3t}
1	7	4	2
2	10	6	4
3	5	2	1
4	11	5	10
5	20	16	3

15.2. Especificación del modelo

Dada la siguiente tabla de datos

Observaciones	Y	X_2	X_3	$\dots$	X_k
1	y_1	x_{21}	x_{31}		x_{k1}
2					
$\vdots$					
t	y_t	x_{2t}	x_{3t}	$\dots$	x_{kt}
$\vdots$					
T	y_T	x_{2T}	x_{3T}	$\dots$	x_{kT}

Si existe una relación lineal entre la variable a explicar Y_t con $k - 1$ variables independientes $X_j, j = 2, \dots, k$ se podría especificar:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t; \quad \forall t = 1, 2, \dots, T \quad [1]$$

En este modelo:

- la variable $Y_t, t = 1, 2, \dots, T$ es la variable endógena
- las variables $X_{jt}, j = 2, \dots, k; t = 1, \dots, T$ son las variables explicativas o exógenas,
- $\boldsymbol{\beta} = [\beta_1 \ \beta_2 \ \dots \ \beta_k]'$ es el vector de coeficientes de regresión,
- $\varepsilon_t, t = 1, \dots, T$ es la perturbación aleatoria,
- T son las unidades de observación temporales o tamaño de la muestra.

Es decir, el modelo se podría expresar como un sistema de ecuaciones, de la siguiente manera

$$\left\{ \begin{array}{l} y_1 = \beta_1 + \beta_2 x_{21} + \beta_3 x_{31} + \cdots + \beta_k x_{k1} + \varepsilon_1 \\ y_2 = \beta_1 + \beta_2 x_{22} + \beta_3 x_{32} + \cdots + \beta_k x_{k2} + \varepsilon_2 \\ y_3 = \beta_1 + \beta_2 x_{23} + \beta_3 x_{33} + \cdots + \beta_k x_{k3} + \varepsilon_3 \\ \vdots \\ y_T = \beta_1 + \beta_2 x_{2T} + \beta_3 x_{3T} + \cdots + \beta_k x_{kT} + \varepsilon_T \end{array} \right. \quad [2]$$

Que es un sistema de T ecuaciones con k incógnitas, por lo que el modelo tiene $T - k$ *grados de libertad*.

Los coeficientes de regresión β_j , $j=1, \dots, k$ se suponen constantes para toda t y recogen el incremento promedio que experimenta la variable endógena cuando se produce un incremento unitario en la j - ésima variable exógena, permaneciendo las demás constantes.

Esta relación es *lineal en los parámetros* y se puede escribir en notación matricial:

$$\underset{\text{Tx1}}{\mathbf{y}} = \underset{\text{Txk}}{\mathbf{X}} \underset{\text{kx1}}{\boldsymbol{\beta}} + \underset{\text{Tx1}}{\boldsymbol{\varepsilon}} \quad [3]$$

Donde cada uno de los elementos se definen según:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_T \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{21} & x_{31} & \dots & x_{k1} \\ 1 & x_{22} & x_{32} & \dots & x_{k2} \\ . & \dots & \dots & \dots & \dots \\ 1 & x_{2T} & x_{3T} & \dots & x_{kT} \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_k \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_T \end{bmatrix}$$

Ejemplo 15.5.a Utilizando la información de la tabla 15.1, la notación es:

$$\mathbf{y} = \begin{bmatrix} 7 \\ 10 \\ 5 \\ 11 \\ 20 \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 & 4 & 2 \\ 1 & 6 & 4 \\ 1 & 2 & 1 \\ 1 & 5 & 10 \\ 1 & 16 & 3 \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix}$$

Supuestos del Modelo Lineal General

El modelo lineal general tiene dos partes: una se denomina sistemática y la otra aleatoria. Esto es,

$$Y_t = \underbrace{\beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt}}_{\text{PARTE SISTEMÁTICA}} + \underbrace{\varepsilon_t}_{\substack{\text{PARTE} \\ \text{ALEATORIA}}} ; \quad \forall t = 1, 2, \dots, T$$

El modelo explica la variabilidad de la variable dependiente Y mediante dos componentes:

- La parte sistemática: $\mathbf{X}\boldsymbol{\beta}$
- La parte aleatoria: $\boldsymbol{\varepsilon}$

La *parte sistemática* del modelo satisface los *supuestos básicos* de:

- 1) *Linealidad*. El comportamiento de la variable dependiente Y se ajusta a los parámetros del modelo de manera lineal para todas las observaciones, $t = 1, \dots, T$.

Sea el vector columna $\mathbf{x}_k$ que contiene las T observaciones de la variable X_k . Este vector representa la k – *ésima* columna de la matriz $\mathbf{X}$ de orden $t \times k$. La primera columna de $\mathbf{X}$ corresponderá a una columna de unos, por lo que β_1 será el término constante del modelo. Si se denomina $\mathbf{y}$ al vector columna que contiene las T observaciones, $y_1, y_2, \dots, y_t$, de la variable endógena Y , $\boldsymbol{\varepsilon}$ al vector columna que contiene las T perturbaciones aleatorias, entonces el modelo puede escribirse como:

$$\mathbf{y} = \beta_1 + \mathbf{x}_2\beta_2 + \dots + \mathbf{x}_k\beta_k + \boldsymbol{\varepsilon}$$

Esto es, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$

Observación. Para evitar posibles confusiones hay que distinguir entre los vectores columna: $\mathbf{x}_k$ y $\mathbf{x}_t$. $\mathbf{x}_k$ es la k – *ésima* columna de $\mathbf{X}$; mientras que, $\mathbf{x}_t$ es un vector columna que es la traspuesta de la fila t – *ésima* ($1 \times k$) de $\mathbf{X}$. Por ejemplo, para referirse a una única unidad de observación se utilizará la ecuación $y_t = \mathbf{x}_t' \boldsymbol{\beta} + \varepsilon_t$; donde $\mathbf{x}_t'$ es la notación que representa la t – *ésima* fila de $\mathbf{X}$.

2) *Variables exógenas no estocásticas*. Las variables explicativas X_j , $j = 2, \dots, k$ se consideran fijas en muestra repetidas y, por lo tanto, la función de distribución de la variable Y - condicionada a los *regresores fijos* X - se puede escribir como $f(Y / X, \beta) = f(Y, \beta)$.

3) *Rango Completo*. El rango de la matriz de variables explicativas es completo por columnas; es decir, $r(\mathbf{X}) = k \leq T$. Este supuesto tiene dos implicaciones. Por un lado, no es posible expresar una columna de la matriz $\mathbf{X}$ como una combinación lineal del resto de las columnas y, por otro, se supone que hay más observaciones que parámetros.

Respecto a su *parte aleatoria*, el modelo lineal general debe cumplir con los siguientes supuestos:

4) *Media Nula*. La esperanza matemática de cada una de las perturbaciones es cero:

$$E(\varepsilon_t) = 0, \quad \forall t = 1, 2, \dots, T \quad [4]$$

De forma matricial, se puede escribir como:

$$E(\boldsymbol{\varepsilon}) = E \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_T \end{bmatrix} = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ E(\varepsilon_3) \\ \vdots \\ E(\varepsilon_T) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

Observación. Aplicar el operador esperanza matemática a una matriz o vector, significa que hay que tomar esperanza matemática de cada uno de los elementos de la matriz o vector en cuestión.

5) La matriz de varianzas y covarianzas del vector de perturbaciones $\boldsymbol{\varepsilon}$ es un escalar: $V(\boldsymbol{\varepsilon}) = \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_T$. Con este supuesto se quiere indicar:

5a) *Homocedasticidad*. La varianza es constante para todas las perturbaciones,

$$E(\varepsilon_t^2) = \sigma_{\boldsymbol{\varepsilon}}^2 \quad \forall t = 1, 2, \dots, T \quad [5]$$

5b) *No autocorrelación*. No existe *correlación* entre las perturbaciones de diferentes períodos,

$$E(\varepsilon_t, \varepsilon_s) = 0 \quad \forall t \neq s \quad [6]$$

La homocedasticidad y no autocorrelación se pueden escribir en forma matricial de la siguiente manera:

$$V(\boldsymbol{\varepsilon}) = \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_T \quad [7]$$

6) *Normalidad*. La distribución de probabilidad del término de perturbación es normal multivariante:

$$\boldsymbol{\varepsilon} \sim N(0, \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_T) \quad [8]$$

El conjunto de supuestos 1) a 6) indican que el comportamiento de la variable Y viene dado por la parte sistemática del modelo y no queda, en la parte aleatoria, algún patrón aprovechable para explicar el comportamiento de la variable dependiente.

Los objetivos, entonces, se centran en hacer inferencia estadística sobre el vector β de parámetros del modelo de regresión y sobre la varianza de la perturbación aleatoria σ_ε^2 en base a la información que proporciona la muestra disponible:

$$(Y_t, X_{2t}, \dots, X_{kt}), t = 1, 2, \dots, T$$

Ejemplo 15.5.b La información de la tabla 15.1 proporciona una muestra de 5 observaciones en el tiempo para las variables Y , X_2 y X_3 , que se utilizará para hacer inferencias sobre el vector paramétrico del modelo.

En este caso dicho vector viene representado por:

$$w = [\beta_j \ \sigma_\varepsilon^2]; \quad \forall j = 1, 2, 3$$

15.3. Estimación por mínimos cuadrados ordinarios (MCO)

Los parámetros desconocidos del modelo vienen dados por el vector de coeficientes de regresión β y la varianza de la perturbación σ_ε^2 .

Ejemplo 15.5.c El siguiente ejemplo ilustra como se obtienen las ecuaciones normales con los datos de la tabla 15.1. Suponga, además, que interesa estimar el modelo:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \varepsilon_t; \quad \forall t = 1, 2, \dots, 5$$

Se puede construir la siguiente tabla de cálculos auxiliares:

Tabla 15.2.

	X_2	X_3	Y	$X_2 Y$	$X_3 Y$	$X_2 X_3$	X_2^2	X_3^2	$\hat{Y}$	$e=Y-\hat{Y}$
1	4	2	7	28	14	8	16	4	7,233	-0,233
2	6	4	10	60	40	24	36	16	9,980	0,020
3	2	1	5	10	5	2	4	1	4,826	0,174
4	5	10	11	55	110	50	25	100	10,986	0,014
5	16	3	20	320	60	48	256	9	19,975	0,025
Σ	33	20	53	473	229	132	337	130	53	0,000

Las ecuaciones normales son:

$$\begin{aligned} 53 &= 5\hat{\beta}_1 + 33\hat{\beta}_2 + 20\hat{\beta}_3 \\ 473 &= 33\hat{\beta}_1 + 337\hat{\beta}_2 + 132\hat{\beta}_3 \\ 229 &= 20\hat{\beta}_1 + 132\hat{\beta}_2 + 130\hat{\beta}_3 \end{aligned}$$

Con solución: $\hat{\beta}_1 = 2.418523$; $\hat{\beta}_2 = 1.033557$ $\hat{\beta}_3 = 0.34$

Por lo tanto el modelo estimado resulta ser:

$$\hat{Y}_t = 2.418523 + 1.033557 X_{2t} + 0.34 X_{3t}$$

Las ecuaciones normales se pueden obtener a través del álgebra matricial, como:

$$(\mathbf{X}' \mathbf{X}) \hat{\boldsymbol{\beta}}_{\text{MCO}} - \mathbf{X}' \mathbf{y} = \mathbf{0}$$

Para obtener estas ecuaciones hay que resolver matricialmente el sistema a minimizar. Para ello, se plantea el siguiente *problema de mínimo*:

La recta de regresión debe pasar por el centro de la nube de puntos. Hay que plantear la minimización de las distancias de esos puntos a la recta. En el plano ($\mathbb{R}^2$) esas distancias son ortogonales a la abscisa -que representa a la variable explicativa-; mientras que, en el espacio de tres dimensiones ($\mathbb{R}^3$) esas distancias son ortogonales al plano -formado por las dos variables explicativas-. En $\mathbb{R}^k$ las distancias son ortogonales al hiperplano formado por las $k-1$ variables explicativas. Esas distancias se denominarán residuos mínimo cuadráticos ordinarios y se simbolizarán por e_t . De esta forma, $e_t = Y_t - \hat{Y}_t$ representa la distancia de cada observación a la recta de regresión a estimar por $\hat{y} = \mathbf{X}\hat{\beta}$, siendo ésta la que se obtendrá a partir de las estimaciones del vector $\hat{\beta}$. Pero, como e_t es una variable desvío, $\sum_{t=1}^T e_t = 0$; por lo que, el problema de mínimo a plantear es el de minimizar la suma de cuadrados de los desvíos.

Esto es

$$\text{Min} \sum_{t=1}^T e_t^2$$

ó, lo que es lo mismo en términos matriciales:

$$\text{Mín} (\mathbf{e}'\mathbf{e}) = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}})$$

Planteado el problema de minimización, se debe resolver algebraicamente la siguiente relación:

$$\text{Min}_{\hat{\beta}} (y - X\hat{\beta})' (y - X\hat{\beta}) = \text{Min}_{\hat{\beta}} [y'y - y'X\hat{\beta} - (X\hat{\beta})'y + (X\hat{\beta})'X\hat{\beta}] \quad [11]$$

Teniendo en cuenta que $y'X\hat{\beta} = (X\hat{\beta})'y$

Entonces

$$[y'y - y'X\hat{\beta} - (X\hat{\beta})'y + (X\hat{\beta})'X\hat{\beta}] = y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}$$

Por lo tanto, la expresión a minimizar será

$$\text{Min}(e'e) = \text{Min} [y'y - 2\hat{\beta}'X'y + \hat{\beta}'X'X\hat{\beta}] \quad [12]$$

Las condiciones de mínimo exigen:

- a) condición de primer orden, derivar respecto a la variable -en este caso $\hat{\beta}$ -, e igualar a cero.
- b) Condición de segundo orden, obtener la segunda derivada y demostrar que es positiva.

Para obtener la condición de primer orden a partir de la expresión [12], hay que tener en cuenta que $\hat{\beta}'X'X\hat{\beta}$ es una *forma cuadrática*; la cual se define como el producto entre un vector, su traspuesto y una matriz simétrica, ordenados de manera que su resultado sea un escalar.

En este caso, la matriz simétrica $\mathbf{X}'\mathbf{X}$ es una matriz de orden $k \times k$, la que se combina con el vector $\hat{\boldsymbol{\beta}}$ de k elementos dando por resultado un escalar. $\mathbf{X}'\mathbf{X}$ se denomina *matriz de la forma cuadrática*.

Algebraicamente

$$\hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 & \cdots & \hat{\beta}_k \end{bmatrix}_{(1 \times k)} \begin{bmatrix} T & \sum_{t=1}^T x_{2t} & \sum_{t=1}^T x_{3t} & \cdots & \sum_{t=1}^T x_{kt} \\ \sum_{t=1}^T x_{2t} & \sum_{t=1}^T x_{2t}^2 & \sum_{t=1}^T x_{2t}x_{3t} & \cdots & \sum_{t=1}^T x_{2t}x_{kt} \\ \sum_{t=1}^T x_{3t} & \sum_{t=1}^T x_{3t}x_{2t} & \sum_{t=1}^T x_{3t}^2 & \cdots & \sum_{t=1}^T x_{3t}x_{kt} \\ \cdots & \cdots & \cdots & \ddots & \cdots \\ \sum_{t=1}^T x_{kt} & \sum_{t=1}^T x_{kt}x_{2t} & \sum_{t=1}^T x_{kt}x_{3t} & \cdots & \sum_{t=1}^T x_{kt}^2 \end{bmatrix}_{(k \times k)} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix}_{(k \times 1)}$$

$$= \begin{bmatrix} \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 & \cdots & \hat{\beta}_k \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}_1 T + \hat{\beta}_2 \sum_{t=1}^T x_{2t} + \hat{\beta}_3 \sum_{t=1}^T x_{3t} + \cdots + \hat{\beta}_k \sum_{t=1}^T x_{kt} \\ \hat{\beta}_1 \sum_{t=1}^T x_{2t} + \hat{\beta}_2 \sum_{t=1}^T x_{2t}^2 + \hat{\beta}_3 \sum_{t=1}^T x_{2t}x_{3t} + \cdots + \hat{\beta}_k \sum_{t=1}^T x_{2t}x_{kt} \\ \hat{\beta}_1 \sum_{t=1}^T x_{3t} + \hat{\beta}_2 \sum_{t=1}^T x_{3t}x_{2t} + \hat{\beta}_3 \sum_{t=1}^T x_{3t}^2 + \cdots + \hat{\beta}_k \sum_{t=1}^T x_{3t}x_{kt} \\ \vdots \\ \hat{\beta}_1 \sum_{t=1}^T x_{kt} + \hat{\beta}_2 \sum_{t=1}^T x_{kt}x_{2t} + \hat{\beta}_3 \sum_{t=1}^T x_{kt}x_{3t} + \cdots + \hat{\beta}_k \sum_{t=1}^T x_{kt}^2 \end{bmatrix} =$$

$$\begin{aligned} &= \hat{\beta}_1 \left[\hat{\beta}_1 T + \hat{\beta}_2 \sum_{t=1}^T x_{2t} + \hat{\beta}_3 \sum_{t=1}^T x_{3t} + \cdots + \hat{\beta}_k \sum_{t=1}^T x_{kt} \right] + \\ &+ \hat{\beta}_2 \left[\hat{\beta}_1 \sum_{t=1}^T x_{2t} + \hat{\beta}_2 \sum_{t=1}^T x_{2t}^2 + \hat{\beta}_3 \sum_{t=1}^T x_{2t}x_{3t} + \cdots + \hat{\beta}_k \sum_{t=1}^T x_{2t}x_{kt} \right] + \\ &+ \hat{\beta}_3 \left[\hat{\beta}_1 \sum_{t=1}^T x_{3t} + \hat{\beta}_2 \sum_{t=1}^T x_{3t}x_{2t} + \hat{\beta}_3 \sum_{t=1}^T x_{3t}^2 + \cdots + \hat{\beta}_k \sum_{t=1}^T x_{3t}x_{kt} \right] + \\ &+ \cdots + \hat{\beta}_k \left[\hat{\beta}_1 \sum_{t=1}^T x_{kt} + \hat{\beta}_2 \sum_{t=1}^T x_{kt}x_{2t} + \hat{\beta}_3 \sum_{t=1}^T x_{kt}x_{3t} + \cdots + \hat{\beta}_k \sum_{t=1}^T x_{kt}^2 \right] = \end{aligned}$$

Reagrupando términos

$$\begin{aligned}\hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta} = & T\hat{\beta}_1^2 + 2\hat{\beta}_1\hat{\beta}_2 \sum_{t=1}^T x_{2t} + 2\hat{\beta}_1\hat{\beta}_3 \sum_{t=1}^T x_{3t} + \cdots + 2\hat{\beta}_1\hat{\beta}_k \sum_{t=1}^T x_{kt} + \\ & + \hat{\beta}_2^2 \sum_{t=1}^T x_{2t}^2 + 2\hat{\beta}_2\hat{\beta}_3 \sum_{t=1}^T x_{2t}x_{3t} + \cdots + 2\hat{\beta}_2\hat{\beta}_k \sum_{t=1}^T x_{2t}x_{kt} + \\ & + \hat{\beta}_3^2 \sum_{t=1}^T x_{3t}^2 + \cdots + 2\hat{\beta}_3\hat{\beta}_k \sum_{t=1}^T x_{3t}x_{kt} + \\ & + \cdots + \hat{\beta}_k^2 \sum_{t=1}^T x_{kt}^2\end{aligned}$$

Esta última expresión es un escalar, que puede derivarse parcialmente con respecto a cada uno de los elementos de $\hat{\beta}$.

El resultado de las derivadas parciales se ordena en forma de vector columna -aunque también podrían ordenarse en forma de vector fila-. No obstante, el requisito importante es la consistencia del tratamiento que debe darse a los vectores y matrices de las derivadas de la función para que sean de orden apropiado para su posterior manipulación.

Entonces, derivando esta expresión respecto de $\hat{\beta}_1$ los primeros k términos, respecto de $\hat{\beta}_2$ los segundos k términos, y así siguiendo... el vector de derivadas parciales es

$$\frac{\partial \hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta}}{\partial \hat{\beta}} = \begin{bmatrix} \frac{\partial \hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta}}{\partial \hat{\beta}_1} \\ \frac{\partial \hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta}}{\partial \hat{\beta}_2} \\ \frac{\partial \hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta}}{\partial \hat{\beta}_3} \\ \vdots \\ \frac{\partial \hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta}}{\partial \hat{\beta}_k} \end{bmatrix} = \begin{bmatrix} 2T\hat{\beta}_1 + 2\hat{\beta}_2 \sum_{t=1}^T x_{2t} + 2\hat{\beta}_3 \sum_{t=1}^T x_{3t} + \cdots + 2\hat{\beta}_k \sum_{t=1}^T x_{kt} \\ 2\hat{\beta}_1 \sum_{t=1}^T x_{2t} + 2\hat{\beta}_2 \sum_{t=1}^T x_{2t}^2 + 2\hat{\beta}_3 \sum_{t=1}^T x_{2t}x_{3t} + \cdots + 2\hat{\beta}_k \sum_{t=1}^T x_{2t}x_{kt} \\ 2\hat{\beta}_1 \sum_{t=1}^T x_{3t} + 2\hat{\beta}_2 \sum_{t=1}^T x_{2t}x_{3t} + 2\hat{\beta}_3 \sum_{t=1}^T x_{3t}^2 + \cdots + 2\hat{\beta}_k \sum_{t=1}^T x_{3t}x_{kt} \\ \vdots \\ 2\hat{\beta}_1 \sum_{t=1}^T x_{kt} + 2\hat{\beta}_2 \sum_{t=1}^T x_{2t}x_{kt} + 2\hat{\beta}_3 \sum_{t=1}^T x_{3t}x_{kt} + \cdots + 2\hat{\beta}_k \sum_{t=1}^T x_{kt}^2 \end{bmatrix} =$$

$$= 2 \cdot \begin{bmatrix} T & \sum_{t=1}^T x_{2t} & \sum_{t=1}^T x_{3t} & \cdots & \sum_{t=1}^T x_{kt} \\ \sum_{t=1}^T x_{2t} & \sum_{t=1}^T x_{2t}^2 & \sum_{t=1}^T x_{2t}x_{3t} & \cdots & \sum_{t=1}^T x_{2t}x_{kt} \\ \sum_{t=1}^T x_{3t} & \sum_{t=1}^T x_{2t}x_{3t} & \sum_{t=1}^T x_{3t}^2 & \cdots & \sum_{t=1}^T x_{3t}x_{kt} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^T x_{kt} & \sum_{t=1}^T x_{2t}x_{kt} & \sum_{t=1}^T x_{3t}x_{kt} & \cdots & \sum_{t=1}^T x_{kt}^2 \end{bmatrix} \cdot \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \quad [13]$$

Este resultado responde a la regla general que establece que “la derivada de una forma cuadrática respecto a cada uno de los elementos del vector de dicha forma, es igual a dos veces el producto de la matriz de la forma cuadrática por el vector de la misma”, en este caso:

$$\frac{\partial \hat{\boldsymbol{\beta}}'(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}}}{\partial \hat{\boldsymbol{\beta}}} = 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \quad [14]$$

[14] es un vector columna de k elementos.

De esta forma, la derivada de la expresión [12] es

$$\frac{\partial \mathbf{e}'\mathbf{e}}{\partial \hat{\boldsymbol{\beta}}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$$

Operando algebraicamente se obtiene la expresión para las *ecuaciones normales*:

$$-\mathbf{X}'\mathbf{y} + \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0} \quad [15]$$

En síntesis, resolver la condición de primer orden en el problema de mínimo da como resultado las ecuaciones normales.

Por otra parte, dado que la matriz $\mathbf{X}$ es de rango completo por columna $[r(\mathbf{X})=k \leq T]$, existe una *solución única* al sistema de ecuaciones normales que es el *estimador mínimo-cuadrático ordinario* del vector de parámetros β :

$$\hat{\beta}_{\text{MCO}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad [16]$$

Donde

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} T & \sum_{t=1}^T X_{2t} & \sum_{t=1}^T X_{3t} & \dots & \sum_{t=1}^T X_{kt} \\ \sum_{t=1}^T X_{2t} & \sum_{t=1}^T X_{2t}^2 & \sum_{t=1}^T X_{2t}X_{3t} & \dots & \sum_{t=1}^T X_{2t}X_{kt} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{t=1}^T X_{kt} & \sum_{t=1}^T X_{kt}X_{2t} & \sum_{t=1}^T X_{kt}X_{3t} & \dots & \sum_{t=1}^T X_{kt}^2 \end{bmatrix}^{-1}; \quad \mathbf{X}'\mathbf{y} = \begin{bmatrix} \sum_{t=1}^T Y_t \\ \sum_{t=1}^T Y_t X_{2t} \\ \vdots \\ \sum_{t=1}^T Y_t X_{kt} \end{bmatrix}$$

Aunque este es el resultado deseado, a partir del cual se obtienen los estimadores mínimos cuadrados ordinarios, aún falta demostrar la condición de *segundo orden* del problema de minimización, esto es:

$$\frac{\partial^2 \mathbf{e}'\mathbf{e}}{\partial \hat{\beta}^2} = 2\mathbf{X}'\mathbf{X} > \mathbf{0} \quad [17]$$

Y esto es así, debido a que la matriz $\mathbf{X}'\mathbf{X}$ es definida positiva.

Si $\hat{\beta}'(\mathbf{X}'\mathbf{X})\hat{\beta} > 0$ para todo $\hat{\beta} \neq \mathbf{0}$, se dice que la forma cuadrática es definida positiva. Una condición necesaria y suficiente para que la forma cuadrática sea definida positiva es que la matriz $\mathbf{X}'\mathbf{X}$ sea definida positiva, esto sucede cuando los menores principales de $\mathbf{X}'\mathbf{X}$

$$|T| \quad \left| \begin{array}{cc} T & \sum_{t=1}^T X_{2t} \\ \sum_{t=1}^T X_{2t} & \sum_{t=1}^T X_{2t}^2 \end{array} \right| \quad \left| \begin{array}{ccc} T & \sum_{t=1}^T X_{2t} & \sum_{t=1}^T X_{3t} \\ \sum_{t=1}^T X_{2t} & \sum_{t=1}^T X_{2t}^2 & \sum_{t=1}^T X_{2t}X_{3t} \\ \sum_{t=1}^T X_{3t} & \sum_{t=1}^T X_{3t}X_{2t} & \sum_{t=1}^T X_{3t}^2 \end{array} \right| \quad \dots \quad |\mathbf{X}'\mathbf{X}|$$

son todos positivos.

Observación. La correspondiente condición necesaria y suficiente para definida negativa es que los menores principales alternen su signo; de forma que todos los menores principales de orden impar sean negativos y todos los de orden par sean positivos. El menor principal de orden k , $|\mathbf{X}'\mathbf{X}|$ será positivo si k es par pero negativo si k es impar. Esto puede expresarse de forma resumida por la desigualdad

$$(-1)^k |\mathbf{X}'\mathbf{X}| > 0$$

Las formas cuadráticas, además de definidas positivas o negativas, pueden ser semidefinidas o indefinidas.

Una forma cuadrática será semidefinida (positiva o negativa) cuando un menor se anule, aun cuando los menores restantes cumplan con el signo correspondiente.

Una forma cuadrática será indefinida cuando no se cumplan las condiciones para que sea definida o semidefinida (positiva o negativa)

Ejemplo 15.5.d Siguiendo con el ejemplo de la tabla 15.1

$$\begin{aligned}
 \hat{\beta}_{MCO} &= \left\{ \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 6 & 2 & 5 & 16 \\ 2 & 4 & 1 & 10 & 3 \end{bmatrix} \begin{bmatrix} 1 & 4 & 2 \\ 1 & 6 & 4 \\ 1 & 2 & 1 \\ 1 & 5 & 10 \\ 1 & 16 & 3 \end{bmatrix} \right\}^{-1} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 4 & 6 & 2 & 5 & 16 \\ 2 & 4 & 1 & 10 & 3 \end{bmatrix} \begin{bmatrix} 7 \\ 10 \\ 5 \\ 11 \\ 20 \end{bmatrix} \\
 &= \begin{bmatrix} 5 & 33 & 20 \\ 33 & 337 & 132 \\ 20 & 132 & 130 \end{bmatrix}^{-1} \begin{bmatrix} 53 \\ 473 \\ 229 \end{bmatrix} \\
 &= \frac{1}{\begin{bmatrix} 5 & 33 & 20 \\ 33 & 337 & 132 \\ 20 & 132 & 130 \end{bmatrix}} \begin{bmatrix} 26386 & -1650 & -2384 \\ -1650 & 250 & 0 \\ -2384 & 0 & 596 \end{bmatrix} \begin{bmatrix} 53 \\ 473 \\ 229 \end{bmatrix} \\
 \hat{\beta}_{MCO} &= \begin{bmatrix} 2.418523 \\ 1.033557 \\ 0.34 \end{bmatrix}
 \end{aligned}$$

El resultado es igual al obtenido al resolver las ecuaciones normales, en el Ejemplo 15.5.c.

De las ecuaciones normales se derivan, entre otras, las siguientes dos *propiedades de la estimación mínimo – cuadrática ordinaria*, la primera de las cuales se ve directamente en la tabla 15.2 del Ejemplo 15.5.c:

$$\sum_{t=1}^T e_t = 0 \quad [18]$$

donde $e_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_{2t} - \dots - \hat{\beta}_k X_{kt}$ son los denominados *residuos mínimo-cuadrático ordinarios*:

$$\sum_{t=1}^T X_{jt} e_t = 0 \quad j = 2, \dots, k \quad [19]$$

es decir, los residuos mínimo cuadráticos ordinarios, son *ortogonales* a todas las variables explicativas del modelo.

Para el caso de $k=2$, $Y_t = \beta_1 + \beta_2 X_{2t} + \varepsilon_t$, la interpretación gráfica de la ortogonalidad de los residuos y de las rectas de regresión es la ilustrada por la Figura 15.1

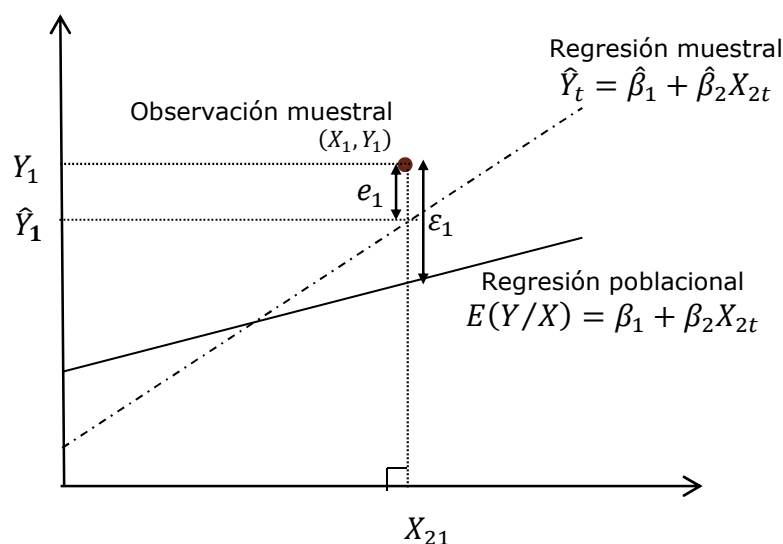


Figura 15.1. Líneas de regresión poblacional y muestral

Para demostrar estas propiedades se considera el vector de residuos mínimos cuadrados ordinarios

$$\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} \Rightarrow \mathbf{y} = \mathbf{e} + \mathbf{X}\hat{\boldsymbol{\beta}} \quad [20]$$

Por [15] $-\mathbf{X}'\mathbf{y} + \mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{0}$, el que puede reexpresarse como

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$$

Reemplazando $\mathbf{y}$ por su igual en [20]

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'(\mathbf{e} + \mathbf{X}\hat{\boldsymbol{\beta}})$$

Realizando los productos convenientemente

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{e} + (\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} \quad [21]$$

Para que la igualdad en [21] se cumpla, debe ocurrir que $\mathbf{X}'\mathbf{e} = \mathbf{0}$; si se desarrolla la expresión

$$(\mathbf{X}'\mathbf{e}) = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_{21} & x_{22} & x_{23} & & x_{2T} \\ x_{31} & x_{32} & x_{33} & & x_{3T} \\ & & & \ddots & \\ x_{k1} & x_{k2} & x_{k3} & & x_{kT} \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \\ e_T \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^T e_t \\ \sum_{t=1}^T x_{2t} e_t \\ \sum_{t=1}^T x_{3t} e_t \\ \vdots \\ \sum_{t=1}^T x_{kt} e_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \quad [22]$$

Como consecuencia de esta propiedad, los residuos del modelo lineal general tienen media aritmética igual a cero (siempre y cuando se incluya el término independiente en la ecuación de regresión). Esto es debido a que el primer elemento del vector de orden $tx1$ es igual a cero, es decir

$$\sum_{t=1}^T e_t = 0 \Rightarrow \bar{e} = \frac{\sum_{t=1}^T e_t}{T} = 0 \quad [23]$$

A su vez, los demás elementos establecen que la correlación muestral entre los residuos y cada variable es cero.

Una vez estimados los coeficientes de regresión, β_j , la recta de regresión muestral,

$$\hat{y} = \mathbf{x}\hat{\beta}$$

permite estimar los valores de la variable endógena Y , dado los valores de las variables exógenas X . Una observación $\hat{Y}_t$ puede ser representada como

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \hat{\beta}_3 X_{3t} + \cdots + \hat{\beta}_k X_{kt}$$

Matricialmente

$$\hat{Y}_t = \mathbf{x}'_t \hat{\beta}$$

Donde $\hat{Y}_t$ es de orden $T \times 1$, $\mathbf{X}_t'$ es de orden $1 \times k$, y $\hat{\beta}$ es de orden $k \times 1$.

$$\text{Esto es, } \hat{Y}_t = [1 \quad X_{2t} \quad X_{3t} \quad \cdots \quad X_{kt}] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad [24]$$

Ejemplo 15.5.e En el ejemplo resulta:

$$\hat{Y}_t = [1 \quad X_{2t} \quad X_{3t}] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

$$\hat{Y}_t = 2.418523 + 1.033557X_{2t} + 0.34X_{3t}$$

Se obtiene, de esta manera, igual resultado que el obtenido en el Ejemplo 15.5.c.

15.4 Propiedades de los estimadores mínimo cuadráticos ordinarios

Los estimadores (MCO), $\hat{\beta}$, bajo los supuestos 1) a 6) son lineales, insesgados y óptimos, en el sentido de tener la mínima varianza dentro de la *clase* de estimadores lineales e insesgados (teorema de Gauss–Markov).

Estimador lineal

La linealidad no requiere de demostración; es evidente, en el cálculo del coeficiente en [16], la relación lineal que une al vector de estimadores con la matriz que contiene los valores observados de las variables.

Estimador insesgado

Para demostrar que el estimador es *Inssegado* se parte de [16]:

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

Utilizando [3]

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

Realizando los productos pertinentes

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

En el primer término del segundo miembro: $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}=\mathbf{I}$, por lo que

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$$

Utilizando el operador esperanza matemática

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon})$$

Por [4], $E(\boldsymbol{\varepsilon})=0$ por lo que

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \quad [25]$$

Estimador óptimo

Un estimador es *Óptimo* cuando tiene mínima varianza. Antes de demostrar esto, se debe hallar la varianza del estimador; la diferencia entre el estimador y su esperanza matemática es igual a

$$\hat{\boldsymbol{\beta}} - E(\hat{\boldsymbol{\beta}}) = \hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} - \boldsymbol{\beta}$$

La varianza del estimador $\mathbf{V}(\hat{\boldsymbol{\beta}})$ se define

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = E\{[\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}][\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}]'\}$$

Ahora bien, al demostrar la propiedad de insesgadez, se obtuvo que

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

De modo que,

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

Entonces:

$$V(\hat{\beta}) = E\left\{(X'X)^{-1}X'\varepsilon[(X'X)^{-1}X'\varepsilon]'\right\}$$

Aplicando las propiedades de matriz traspuesta:

$$= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}]$$

Introduciendo el operador esperanza matemática

$$= (X'X)^{-1}X'E(\varepsilon\varepsilon')X(X'X)^{-1}$$

Por lo establecido en [7], $V(\epsilon) = E(\epsilon'\epsilon) = \sigma_\epsilon^2 \mathbf{I}_T$; al reemplazarlo en la expresión anterior

$$= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \sigma_\epsilon^2 \mathbf{I}_T \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

σ_ϵ^2 es una constante, por lo que premultiplica al resto de la expresión

$$= \sigma_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \mathbf{I}_T \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$$

Simplificando en la expresión anterior $\mathbf{X}' \mathbf{I}_T \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}$, la varianza del estimador es el producto entre la varianza del término de perturbación y la matriz inversa de $(\mathbf{X}'\mathbf{X})$

$$V(\hat{\beta}) = \sigma_\epsilon^2 (\mathbf{X}'\mathbf{X})^{-1} \quad [26]$$

Para demostrar que esta varianza es mínima, se supone otro estimador

$$\beta^* = [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{P}] \mathbf{y} \quad [27]$$

Donde $\mathbf{P}$ es cualquier matriz de orden $(\mathbf{k} \times \mathbf{T})$ que en caso de anularse hace que $\beta^* = \hat{\beta}$.

Ahora, reemplazando [3] en [27]

$$\beta^* = [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' + \mathbf{P}](\mathbf{X}\beta + \epsilon) = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \epsilon + \mathbf{P}\mathbf{X}\beta + \mathbf{P}\epsilon$$

Al tomar esperanza matemática

$$E(\beta^*) = \beta + (X'X)^{-1}X'E(\varepsilon) + PX\beta + PE(\varepsilon)$$

Aplicando lo establecido en [4]

$$E(\beta^*) = \beta + PX\beta$$

Si $PX = 0$

$$E(\beta^*) = \beta$$

Lo que significa que β^* es un estimador insesgado

El cálculo de la varianza de β^* es

$$V(\beta^*) = E\{[\beta^* - \beta][\beta^* - \beta]'\}$$

Donde

$$\beta^* - \beta = \beta + (X'X)^{-1}X'\varepsilon + \underbrace{PX\beta}_0 + P\varepsilon - \beta$$

Manteniendo la restricción $PX = 0$

$$\beta^* - \beta = (X'X)^{-1}X'\varepsilon + P\varepsilon$$

Reordenando

$$\beta^* - \beta = [(X'X)^{-1}X' + P]\varepsilon$$

Reemplazando $\beta^* - \beta$ en $V(\beta^*)$, se tiene que

$$V(\beta^*) = E \left\{ \left[(X'X)^{-1}X' + P \right] \varepsilon \varepsilon' \left[X(X'X)^{-1} + P' \right] \right\}$$

Introduciendo el operador esperanza

$$V(\beta^*) = [(X'X)^{-1}X' + P] E(\varepsilon \varepsilon') [X(X'X)^{-1} + P']$$

Por [7], $V(\varepsilon) = E(\varepsilon \varepsilon') = \sigma_\varepsilon^2 I_T$

$$V(\beta^*) = [(X'X)^{-1}X' + P] \sigma_\varepsilon^2 I_T [X(X'X)^{-1} + P']$$

Teniendo en cuenta que σ_ε^2 es constante y realizando los productos:

$$V(\beta^*) = \sigma_\varepsilon^2 [(X'X)^{-1}X'X(X'X)^{-1} + PX(X'X)^{-1} + (X'X)^{-1}X'P' + PP']$$

Introduciendo la restricción $PX = 0$, que da lugar a que $X'P' = 0$, y operando algebraicamente, la expresión anterior se reduce a:

$$V(\beta^*) = \sigma_\varepsilon^2 [(X'X)^{-1} + PP']$$

Por lo tanto,

$$V(\beta^*) = \sigma_\varepsilon^2 [(X'X)^{-1} + PP'] > V(\hat{\beta}) \quad [28]$$

La diferencia entre las varianzas de $\hat{\beta}$ y β^* es PP' , lo que hace que $V(\beta^*) > V(\hat{\beta})$

A modo de conclusión, en el modelo $Y = X\beta + \varepsilon$ los β parámetros desconocidos se reemplazan por los $\hat{\beta}$ estimadores mínimos cuadrados ordinarios y como consecuencia se tiene que:

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

Donde cada $\hat{\beta}_j, \forall j = 1, 2, \dots, k$ es el mejor estimador lineal insesgado y óptimo (ELIO) de β_j .

Por lo tanto, dado

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t$$

el *ELIO* de cualquier combinación lineal de los β parámetros es esa misma combinación lineal de los $\hat{\beta}$ estimadores

El *ELIO* de $E(Y_t)$ es $\hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \dots + \hat{\beta}_k X_{kt}$

Varianza del estimador

La varianza de las perturbaciones σ_ε^2 , se puede estimar mediante la expresión:

$$S^2 = \frac{\mathbf{e}'\mathbf{e}}{T-k} \quad [29]$$

Ejemplo 15.5.f El valor de las varianzas de las perturbaciones para el ejemplo que se está desarrollando:

$$S^2 = \frac{\mathbf{e}'\mathbf{e}}{T-k} = \frac{[-0.233 \quad 0.020 \quad 0.174 \quad 0.014 \quad 0.025] \begin{bmatrix} -0.233 \\ 0.020 \\ 0.174 \\ 0.014 \\ 0.025 \end{bmatrix}}{5-3}$$

$$S^2 = \frac{\mathbf{e}'\mathbf{e}}{T-k} = \frac{0.08577}{5-3} = 0.04289$$

Este estimador es insesgado bajo los supuestos 1) a 6)

Para conocer la *precisión* con que se estiman los parámetros, es necesario derivar la matriz de varianzas y covarianzas de los estimadores que, bajo los supuestos habituales, es de la forma

$$\mathbf{v}(\hat{\boldsymbol{\beta}}) = \sigma_{\varepsilon}^2 (\mathbf{X}'\mathbf{X})^{-1} \quad [30]$$

Un estimador insesgado de $\mathbf{v}(\hat{\boldsymbol{\beta}})$, se puede obtener sustituyendo en [30] la varianza de las perturbaciones por su estimador insesgado: S^2

$$\hat{\mathbf{v}}(\hat{\boldsymbol{\beta}}) = S^2 (\mathbf{X}'\mathbf{X})^{-1} \quad [31]$$

Ejemplo 15.5.g El cálculo para los datos de la tabla 15.1 es

$$\begin{aligned}
 \hat{V}(\hat{\beta}) &= 0.04289 \begin{bmatrix} 5 & 33 & 20 \\ 33 & 120 & 39 \\ 20 & 39 & 40 \end{bmatrix}^{-1} = 0.04289 \begin{bmatrix} 0,8854 & -0,0554 & -0,08 \\ -0,0554 & 0,0084 & 0 \\ -0,08 & 0 & 0,02 \end{bmatrix} \\
 &= \begin{bmatrix} 0,03797 & -0,00237 & -0,00343 \\ -0,00237 & -0,00036 & 0 \\ -0,00343 & 0 & 0,00086 \end{bmatrix} \\
 &= \begin{bmatrix} \hat{V}(\hat{\beta}_1) & Cov(\hat{\beta}_1\hat{\beta}_2) & Cov(\hat{\beta}_1\hat{\beta}_3) \\ Cov(\hat{\beta}_2\hat{\beta}_1) & \hat{V}(\hat{\beta}_2) & Cov(\hat{\beta}_2\hat{\beta}_3) \\ Cov(\hat{\beta}_3\hat{\beta}_1) & Cov(\hat{\beta}_3\hat{\beta}_2) & \hat{V}(\hat{\beta}_3) \end{bmatrix}
 \end{aligned}$$

Que es la matriz de varianzas y covarianzas de los estimadores para los datos del ejemplo.

La estimación anterior es posible demostrarla a partir de la suma de cuadrado de los residuos (*SCR*). Esta suma es un escalar que se puede calcular a partir de

$$SCR = \sum_{t=1}^T e_t^2 = \mathbf{e}'\mathbf{e} \quad [32]$$

Por [20], $\mathbf{e} = \mathbf{y} - \mathbf{X}\hat{\beta}$

Utilizando el resultado de [16]

$$\mathbf{e} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

Reagrupando términos en torno a $\mathbf{y}$

$$\mathbf{e} = \underbrace{[\mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}']}_M \mathbf{y}$$

El coeficiente de $\mathbf{y}$ se denomina matriz $\mathbf{M}$

$$\mathbf{e} = \mathbf{M}\mathbf{y}$$

$\mathbf{M}$ es una matriz de orden $\mathbf{T} \times \mathbf{T}$ que posee propiedades interesantes a los efectos de su posterior tratamiento; es idempotente ($\mathbf{M}^2 = \mathbf{M}$), es simétrica ($\mathbf{M}' = \mathbf{M}$) y $\mathbf{M}\mathbf{X} = \mathbf{0}$.

Reemplazando $\mathbf{y}$ por su igual en [3]

$$\mathbf{e} = \mathbf{M}\mathbf{y} = \mathbf{M}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \quad [33]$$

resolviendo

$$\mathbf{e} = \mathbf{M}\mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\varepsilon}$$

Aplicando las propiedades de la matriz $\mathbf{M}$

$$\mathbf{e} = \mathbf{M}\boldsymbol{\varepsilon} \quad [34]$$

Por lo que la suma de cuadrados de los errores será

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}'\mathbf{M}\boldsymbol{\varepsilon}$$

Aplicando las propiedades de simetría e idempotencia se obtiene un escalar,

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}^2\boldsymbol{\varepsilon}$$

$$\mathbf{e}'\mathbf{e} = \boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon} \quad [35]$$

Si al escalar definido en [35] se le aplica el operador esperanza matemática, se obtiene

$$E(\mathbf{e}'\mathbf{e}) = E(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon})$$

si a esta igualdad le aplicamos la traza

$$E(\mathbf{e}'\mathbf{e}) = E[\text{tr}(\boldsymbol{\varepsilon}'\mathbf{M}\boldsymbol{\varepsilon})]$$

Observación. Sea $\mathbf{A}$, una matriz de $m \times m$, la traza de $\mathbf{A}$ se define como la suma de los elementos de la diagonal principal

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^m a_{ii} = a_{11} + a_{22} + \cdots + a_{mm}$$

En particular,

$$\text{tr}(\mathbf{I}_T) = T$$

La traza de una matriz tiene las siguientes propiedades:

Sean $\mathbf{A}$ y $\mathbf{B}$ matrices de $m \times m$

$$\text{tr}(k\mathbf{A}) = k \text{tr}(\mathbf{A})$$

$$\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$$

$$\text{tr}(\mathbf{A}') = \text{tr}(\mathbf{A})$$

Si $\mathbf{A}$ y $\mathbf{B}$ son matrices $m \times n$ y $n \times m$, respectivamente, entonces

$$\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$$

por lo que

$$E(\mathbf{e}'\mathbf{e}) = E[\text{tr}(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')]]$$

Pero la traza de un escalar es igual al mismo escalar

$$E(\mathbf{e}'\mathbf{e}) = \text{tr}E(\mathbf{M}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$$

Pero $\mathbf{M}$ depende de $\mathbf{X}$ que es no estocástica; es decir, $\mathbf{M}$ es una constante, por lo que

$$E(\mathbf{e}'\mathbf{e}) = \text{tr}[\mathbf{M}E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')]$$

Nuevamente, por [7] $V(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_T$

$$E(\mathbf{e}'\mathbf{e}) = \text{tr}(\mathbf{M}\sigma_{\boldsymbol{\varepsilon}}^2 \mathbf{I}_T)$$

De modo que

$$E(\mathbf{e}'\mathbf{e}) = \sigma_{\boldsymbol{\varepsilon}}^2 \text{tr}\mathbf{M} \quad [36]$$

Pero la traza de la matriz $\mathbf{M}$ es

$$\begin{aligned} \text{tr}\mathbf{M} &= \text{tr}[\mathbf{I}_T] - \text{tr}[\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'] \\ &= \text{tr}[\mathbf{I}_T] - \text{tr}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}] \end{aligned}$$

Esto es así, ya que $\text{tr}\mathbf{AB} = \text{tr}\mathbf{BA}$ (en este caso $\mathbf{A} = \mathbf{X}$ y $\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$), de modo que:

$$= \text{tr}[\mathbf{I}_T] - \text{tr}[\mathbf{I}_k] = T - k$$

Por consiguiente, en [36]

$$E(\mathbf{e}'\mathbf{e}) = \sigma_{\epsilon}^2(T - k) \quad [37]$$

De donde se observa inmediatamente que: $S^2 = \frac{\mathbf{e}'\mathbf{e}}{T - k}$, es un estimador insesgado de σ_{ϵ}^2 .

Con este último punto se concluye con la tarea de especificar y estimar un modelo por el método de mínimos cuadrados ordinarios, con todas las propiedades que hacen a esos estimadores lineales, insesgados y óptimos (ELIO).

15.5. Normalidad de la perturbación aleatoria

Se ha admitido, aunque implícitamente, que los datos con que se trabaja provienen de *muestras finitas*. Esto es cierto en la mayoría de los casos, aunque para sostenerlo son necesarios supuestos bastante fuertes, tales como regresores no estocásticos y distribución normal de las perturbaciones aleatorias. Vale decir aquí que, además de mínimos cuadrados ordinarios existen otros

métodos para estimar el vector de parámetros β . Por ejemplo, se podría aplicar una *regresión por cuantiles* (separando los valores más altos y más pequeños de las variables explicativas) o una *regresión ortogonal* (minimizando la distancia ortogonal a la recta de ajuste y no la distancia ortogonal respecto a las variables explicativas). La cuestión de cuál estimador hay que elegir normalmente se basa en las propiedades estadísticas de los candidatos, tales como insesgadez, eficiencia y precisión. Estos, a su vez, dependen también de la distribución que se supone que producen los datos. Es interesante el hecho de que un buen número de propiedades deseables pueden obtenerse para el estimador (MCO), incluso sin especificar una distribución particular para las perturbaciones aleatorias en la regresión. Sin embargo, se admite -a los efectos de ampliar la discusión- que las perturbaciones siguen una distribución normal. Esto es, se incluye el supuesto adicional de *Normalidad* y se incorporan algunas propiedades asintóticas.

En forma alternativa se podrían calcular los *estimadores máximo verosímiles*, de los parámetros del modelo; es decir, aquellos que son más probables dada la distribución de los datos muestrales y su implicación sobre la función de densidad conjunta.

Para facilitar una posterior comprensión se introduce a continuación resultados estadísticos básicos en forma matricial. En el modelo lineal general ε representa un vector columna de T variables aleatorias, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T$. El valor esperado de cada variable es $E(\varepsilon_t)$ $t = 1, 2, \dots, T$.

Agrupando estos valores esperados en un vector, se obtiene

$$E(\boldsymbol{\varepsilon}) = \begin{bmatrix} E(\varepsilon_1) \\ E(\varepsilon_2) \\ \vdots \\ E(\varepsilon_T) \end{bmatrix} \quad [38]$$

La aplicación del operador E (esperanza) al vector $\boldsymbol{\varepsilon}$ significa que E se aplica a cada elemento del vector.

La varianza de ε_t es, por definición de varianza $V(\varepsilon_t) = E[\varepsilon_t - E(\varepsilon_t)]^2$.

La covarianza entre ε_t y ε_s es $E(\varepsilon_t, \varepsilon_s) = E\{[\varepsilon_t - E(\varepsilon_t)] * [\varepsilon_s - E(\varepsilon_s)]\}$.

Si se define el vector

$$\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon}) = \begin{bmatrix} E(\varepsilon_1 - E(\varepsilon_1)) \\ E(\varepsilon_2 - E(\varepsilon_2)) \\ \vdots \\ E(\varepsilon_T - E(\varepsilon_T)) \end{bmatrix}$$

y se toma $E\{[\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})][\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})]'\}$ queda $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}')$, ya que por el supuesto del modelo $E(\boldsymbol{\varepsilon}) = \mathbf{0}$

Esto es igual a decir

$$V(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = E \left[\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix} [\varepsilon_1 \quad \varepsilon_2 \quad \cdots \quad \varepsilon_T] \right]$$

$$= E \begin{bmatrix} \varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \varepsilon_1\varepsilon_3 & \cdots & \varepsilon_1\varepsilon_T \\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \varepsilon_2\varepsilon_3 & \cdots & \varepsilon_2\varepsilon_T \\ \varepsilon_3\varepsilon_1 & \varepsilon_3\varepsilon_2 & \varepsilon_3^2 & \cdots & \varepsilon_3\varepsilon_T \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \varepsilon_T\varepsilon_1 & \varepsilon_T\varepsilon_2 & \varepsilon_T\varepsilon_3 & \cdots & \varepsilon_T^2 \end{bmatrix} = \begin{bmatrix} E(\varepsilon_1^2) & E(\varepsilon_1\varepsilon_2) & E(\varepsilon_1\varepsilon_3) & \cdots & E(\varepsilon_1\varepsilon_T) \\ E(\varepsilon_2\varepsilon_1) & E(\varepsilon_2^2) & E(\varepsilon_2\varepsilon_3) & \cdots & E(\varepsilon_2\varepsilon_T) \\ E(\varepsilon_3\varepsilon_1) & E(\varepsilon_3\varepsilon_2) & E(\varepsilon_3^2) & \cdots & E(\varepsilon_3\varepsilon_T) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_T\varepsilon_1) & E(\varepsilon_T\varepsilon_2) & E(\varepsilon_T\varepsilon_3) & \cdots & E(\varepsilon_T^2) \end{bmatrix} =$$

Teniendo en cuenta las definiciones dadas, los elementos de esta matriz son las varianzas y covarianzas de las variables ε_t , la cual se puede representar como:

$$= \begin{bmatrix} V(\varepsilon_1) & \text{Cov}(\varepsilon_1\varepsilon_2) & \text{Cov}(\varepsilon_1\varepsilon_3) & \cdots & \text{Cov}(\varepsilon_1\varepsilon_T) \\ \text{Cov}(\varepsilon_2\varepsilon_1) & V(\varepsilon_2) & \text{Cov}(\varepsilon_2\varepsilon_3) & \cdots & \text{Cov}(\varepsilon_2\varepsilon_T) \\ \text{Cov}(\varepsilon_3\varepsilon_1) & \text{Cov}(\varepsilon_3\varepsilon_2) & V(\varepsilon_3) & \cdots & \text{Cov}(\varepsilon_3\varepsilon_T) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(\varepsilon_T\varepsilon_1) & \text{Cov}(\varepsilon_T\varepsilon_2) & \text{Cov}(\varepsilon_T\varepsilon_3) & \cdots & V(\varepsilon_T) \end{bmatrix} =$$

Las varianzas son los elementos de la diagonal principal y las covarianzas aquellos que se encuentran fuera de ésta diagonal. Teniendo en cuenta el supuesto 5 (5.a y 5.b) la matriz es igual a:

$$\begin{bmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & \sigma_\varepsilon^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_\varepsilon^2 \end{bmatrix} = \sigma_\varepsilon^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Esta matriz se conoce como matriz de covarianzas y se simboliza como

$$\Sigma = V(\varepsilon) = E(\varepsilon\varepsilon') = \sigma_\varepsilon^2 \mathbf{I}_T \quad [39]$$

Está claro que Σ es simétrica (esto es, $\Sigma = \Sigma'$). Además, Σ es definida positiva ya que todos sus menores principales son positivos.

Para que Σ sea definida positiva las ε deben ser linealmente independientes. Si se define una variable aleatoria escalar Y como una combinación lineal de las ε

$$Y = [\varepsilon - E(\varepsilon)]'c \quad [40]$$

donde c es un vector columna arbitrario de t elementos no todos nulos.

Elevando [40] al cuadrado

$$Y^2 = c'[\varepsilon - E(\varepsilon)][\varepsilon - E(\varepsilon)]'c$$

esto es, por ser un escalar de argumento vectorial, su cuadrado se obtiene premultiplicando por su transpuesta.

Aplicando el operador esperanza, se obtiene

$$E(Y^2) = E\{c'[\varepsilon - E(\varepsilon)][\varepsilon - E(\varepsilon)]'c\} = c'E\{[\varepsilon - E(\varepsilon)][\varepsilon - E(\varepsilon)]'\}c$$

por ser c un vector de elementos constantes y ε un vector de variables aleatorias, entonces

$$E(Y^2) = c'\Sigma c \quad [41]$$

Puesto que Y es una variable aleatoria escalar se cumplirá que: $E(Y^2) \geq 0$ de esta forma, $c'\Sigma c \geq 0$ y Σ es semidefinida positiva.

Como se ve, $E(Y^2)$ puede asumir un valor nulo o un valor mayor que cero. Si asume un valor nulo, se tiene $E(Y^2) = 0$ lo que implica que $Y = 0$.

Observación. La esperanza matemática de una constante s es igual a la constante s , por lo tanto, la esperanza matemática del cuadrado de una constante es el mismo cuadrado:

$$E(s)=s \Rightarrow E(s^2)=s^2$$

Si la constante es 0, la esperanza matemática es 0. De esta forma:

$$E(Y^2)=0 \Rightarrow Y^2=0 \text{ por lo tanto, } Y=0$$

Entonces, $[\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})]' \mathbf{c} = \mathbf{0}$; pero como $\mathbf{c}$ no es un vector nulo la única posibilidad es que $[\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon})] = \mathbf{0}$, lo cual significa que las desviaciones de $\boldsymbol{\varepsilon}$ con respecto a su media, esto es $[\varepsilon_1 - E(\varepsilon_1)], [\varepsilon_2 - E(\varepsilon_2)], \dots, [\varepsilon_T - E(\varepsilon_T)]$, son linealmente dependientes.

De modo que $\boldsymbol{\Sigma}$ es definida positiva si y solo si entre las ε no existe dependencia lineal.

Observación. Si para un conjunto de parámetros λ , no todos nulos, pertenecientes a un campo numérico F tenemos que se cumple la siguiente combinación lineal $\lambda_1 \mathbf{a}_1 + \lambda_2 \mathbf{a}_2 + \dots + \lambda_n \mathbf{a}_n = \mathbf{0}$ se dice que los vectores $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ son linealmente dependientes dentro de F , salvo que la igualdad se cumpla solo y solo si todos los λ_i ($i=1, 2, \dots, n$) son iguales a cero. Entonces se dice que los vectores son LI. Esta definición se aplica también cuando el número de vectores es uno, de modo tal que un único vector $\mathbf{a}_1$ es independiente si $\mathbf{a}_1 \neq \mathbf{0}$ y dependiente si $\mathbf{a}_1 = \mathbf{0}$, es decir, es el vector nulo. En el caso que nosotros analizamos se

cumple esta última condición ya que cada variable desvío es nula y conforma un vector nulo: $\mathbf{x} = (\mathbf{x} - \boldsymbol{\mu})' = \mathbf{0}$; esto es, $[x_1, x_2, \dots, x_n] = [(X_1 - \mu_1), (X_2 - \mu_2), \dots, (X_n - \mu_n)] = [0, 0, \dots, 0]$ donde x_i es la variable desvío.

Observación. Dadas t variables aleatorias cualesquiera, con alguna función de densidad de probabilidad multivariante

$$p(\mathbf{x}) = p(X_1, X_2, \dots, X_t)$$

Con $E(\mathbf{x}) = \boldsymbol{\mu}$ y $V(\mathbf{x}) = \boldsymbol{\Sigma}$

La función de densidad de probabilidad más importante es la normal multivariante que, al igual que la univariante, queda especificada una vez que se conoce su media y su varianza. En este caso se puede especificar en términos del vector de medias $\boldsymbol{\mu}$ y de su matriz de varianzas y covarianzas $\boldsymbol{\Sigma}$. De este modo la fórmula es:

$$p(\mathbf{X}) = \frac{1}{(2\pi)^{T/2} |\boldsymbol{\Sigma}|^{1/2}} e^{\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]} \quad [42]$$

donde:

$\boldsymbol{\Sigma}$ es una matriz simétrica, definida positiva, cuyos elementos σ_{ts} son parámetros

$\boldsymbol{\mu}$ es un vector $t \times 1$, cuyos elementos μ_t son parámetros.

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_t \end{bmatrix}$$

Una forma compacta de escribir [42] es

$$\mathbf{x} \sim N(\boldsymbol{\mu}; \boldsymbol{\Sigma})$$

es decir, el vector $\mathbf{x}$ de variables X_t se distribuye según una ley normal multivariante con vector de medias $\boldsymbol{\mu}$ y matriz de varianzas y covarianzas $\boldsymbol{\Sigma}$.

Casos especiales

a) Cuando $t=1$,

$$\Sigma = \sigma_{11} = E[(X_1 - \mu_1)(X_1 - \mu_1)] = E[(X_1 - \mu_1)^2] = \sigma_1^2 = \sigma^2 \quad [43]$$

y [42] se transforma en

$$p(X) = \frac{1}{(2\pi)^{1/2} \sigma} e^{\left[-\frac{1}{2\sigma^2}(x-\mu)^2\right]} \quad [44]$$

que es la conocida función de densidad para una normal univariante.

b) La forma cuadrática de la normal multivariante se define como

$$Q = (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \quad [45]$$

Es una forma cuadrática en los elementos $X_t - \mu_t$, y puede escribirse así:

$$Q = \sum_{s=1}^S \sum_{t=1}^T \frac{(X_t - \mu_t)(X_s - \mu_s)}{\sigma_{ts}} \quad [46]$$

Como quedó demostrado la matriz de la forma cuadrática, $\boldsymbol{\Sigma}$, es definida positiva por lo que la forma cuadrática también lo es. Un resultado inmediato de esto es que $p(\mathbf{x}) > 0$, puesto

que el determinante de una matriz definida positiva es positivo, $|\Sigma| > 0$. Esto basta para probar que [42] satisface una de las propiedades que la califican como función de densidad. La otra propiedad que deberíamos probar es que $\int \dots \int p(\mathbf{x}) = 1$. Cuestión que se cumple pero que no demostraremos aquí.

- c) Un caso especialmente importante de [42] se da cuando todas las X tienen la misma varianza σ^2 y no están correlacionadas entre sí (lo que es lo mismo decir que son estadísticamente independientes). Para que esto ocurra Σ debe ser una matriz diagonal, como es el caso de la matriz de varianzas y covarianzas de la variable ϵ

$$\Sigma = \sigma_{\epsilon}^2 \mathbf{I}_T = \begin{bmatrix} \sigma_{\epsilon}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\epsilon}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{\epsilon}^2 \end{bmatrix} \quad [47]$$

Esta matriz tiene las siguientes particularidades

$$|\Sigma| = \sigma_{\epsilon}^{2T}$$

$$|\Sigma|^{1/2} = (\sigma_{\epsilon}^{2T})^{1/2} = (\sigma_{\epsilon}^2)^{T/2}$$

$$|\Sigma|^{-1} = \frac{1}{\sigma_{\epsilon}^2} \mathbf{I}$$

con lo que

$$p(\epsilon) = \frac{1}{(2\pi\sigma^2)^{t/2}} e^{\left[-\frac{1}{2\sigma_{\epsilon}^2} \epsilon' \epsilon\right]} \quad [48]$$

La ecuación [48] se puede factorizar de la siguiente forma:

$$\begin{aligned}
 p(\boldsymbol{\varepsilon}) = p(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T) &= \prod_{t=1}^T \left\{ \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} e^{\frac{1}{2\sigma^2}[\varepsilon_t - E(\varepsilon_t)]} \right\} \\
 &= p(\varepsilon_1)p(\varepsilon_2) \dots p(\varepsilon_T) \quad [49]
 \end{aligned}$$

de modo que la densidad multivariante es el producto de cada una de las densidades marginales; es decir, las ε se distribuyen independientemente unas de otras. Este resultado es de gran importancia; si los coeficientes de correlación entre variables que se distribuyen normalmente son cero entonces las variables son estadísticamente independientes.

No se puede generalizar este resultado a cualquier tipo de distribución y deberá tenerse presente que las correlaciones que deben ser cero son las poblacionales y no las muestrales.

Ejemplo 15.6 Dada una matriz de 3x3

$$B = \begin{bmatrix} \sigma_{11} & 0 & 0 \\ 0 & \sigma_{22} & 0 \\ 0 & 0 & \sigma_{33} \end{bmatrix}$$

Donde:

3 es el número de variables

$$\sigma_{11} = \sigma_{22} = \sigma_{33} = \sigma^2$$

$\sigma_{ij} = 0, \forall i \neq j$, lo cual indica que el coeficiente de correlación ρ_{ij} es cero cuando $i \neq j$.

El determinante es

$$|B| = \sigma^{2*3} \Rightarrow |B| = \sigma_{11}\sigma_{22}\sigma_{33} = \sigma^2\sigma^2\sigma^2 = \sigma^{2*3} = \sigma^6$$

La inversa es

$$B^{-1} = \frac{1}{\sigma^2} I$$

Para el cálculo se utilizará el método de la matriz adjunta, por el cual

$$B^{-1} = \frac{1}{|B|} \text{Adj}(B)$$

$$\text{Adj}(B) = [\text{Cof}(B)]' = [-1^{i+j} M_{i+j}^B]' = \begin{bmatrix} (\sigma_\varepsilon^2)^2 & 0 & 0 \\ 0 & (\sigma_\varepsilon^2)^2 & 0 \\ 0 & 0 & (\sigma_\varepsilon^2)^2 \end{bmatrix} = \sigma_\varepsilon^2 I$$

De modo que:

$$B^{-1} = \frac{1}{\sigma_\varepsilon^6} \sigma_\varepsilon^4 I = \frac{1}{\sigma_\varepsilon^2} I$$

Quedan demostradas las particularidades que tiene la matriz Σ

Recordando la hipótesis de normalidad en la distribución de la perturbación aleatoria dada en [6]

$$\varepsilon \sim N(0, \sigma_\varepsilon^2 I_T)$$

Es decir, el vector de perturbaciones aleatorias $\boldsymbol{\varepsilon}$ tiene una distribución normal multivariante, dada por [48], donde:

$\sigma_{\varepsilon}^2 \mathbf{I}_T$ es una matriz escalar, definida positiva, cuyos elementos σ_{ε}^2 son constantes.

Si se recuerda la definición de la distribución normal multivariante, se observa que en este caso todas las variables tienen la misma varianza σ^2 y no están correlacionadas entre sí (lo que es lo mismo decir que son estadísticamente independientes). Para que esto ocurra $\boldsymbol{\Sigma}$ debe ser una matriz diagonal, esto es

$$\boldsymbol{\Sigma} = \sigma^2 \mathbf{I} = \begin{bmatrix} \sigma_{11} & 0 & \cdots & 0 \\ 0 & \sigma_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_{TT} \end{bmatrix}$$

Con las siguientes propiedades

- $\sigma_{11} = \sigma_{22} = \cdots = \sigma_{TT} = \sigma^2$
- $\sigma_{ij} = 0, \forall i \neq j$. Esto ocurre si y solamente si el coeficiente de correlación ρ_{ij} es cero cuando $i \neq j$.
- $|\boldsymbol{\Sigma}| = \sigma^{2T}$; $|\boldsymbol{\Sigma}|^{1/2} = (\sigma^{2T})^{1/2} = (\sigma^2)^{T/2}$ y $\boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma^2} \mathbf{I}$, con lo que se obtiene la función mostrada.

Por tanto, los elementos del vector $\boldsymbol{\varepsilon}$, se distribuyen independiente y conjuntamente según una ley normal multivariante con vector de medias $\mathbf{0}$ y matriz de covarianzas $\sigma_{\varepsilon}^2 \mathbf{I}_T$.

15.6. Criterio de máxima verosimilitud

En general, las propiedades asintóticas del estimador máximo-verosímil son muy atractivas en casos en los que es imposible encontrar estimadores con buenas propiedades para muestras finitas, situación esta que se produce frecuentemente en la práctica.

Para ello, si se supone que las perturbaciones aleatorias siguen una distribución normal multivariante como la expuesta en [48]

$$f(\boldsymbol{\varepsilon}) = (2\pi)^{-T/2} (\sigma_{\varepsilon}^2)^{-T/2} \exp\left[-\frac{1}{2\sigma_{\varepsilon}^2} \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}\right]$$

La función de verosimilitud, para los valores muestrales, expresando $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ y denominando $\hat{\boldsymbol{\beta}}_{MV}$ al vector de estimadores máximo verosímiles, es

$$L = (2\pi\sigma_{\varepsilon}^2)^{-T/2} \exp\left[-\frac{1}{2\sigma_{\varepsilon}^2} \sum_{t=1}^T (y_t - \mathbf{x}_t'\boldsymbol{\beta})^2\right]$$

$$L = (2\pi\sigma_{\varepsilon}^2)^{-T/2} \exp\left[-\frac{1}{2\sigma_{\varepsilon}^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right] \quad [50]$$

Observación. La transformación $\varepsilon_t = y_t - \mathbf{x}_t'\boldsymbol{\beta}$ es posible ya que el jacobino para cada observación, $\partial\varepsilon_t/\partial y_t$ es igual a la unidad. Ante una transformación de variables se aplica la solución estadística de cambio de variable.

Mediante un cambio de variable se puede resolver, totalmente o en parte, un buen número de problemas importantes en la teoría estadística. Los cambios de variables pueden ser simples cambios de localización o escala o pueden ser transformaciones ortogonales.

En este caso se tiene una variable aleatoria $\boldsymbol{\varepsilon}$ con comportamiento aleatorio conocido, esto es, con densidad $f(\boldsymbol{\varepsilon})$ conocida, y se necesita determinar el comportamiento aleatorio o la densidad $g(\mathbf{y})$, de una variable aleatoria $\mathbf{y}$ cuya relación con $\boldsymbol{\varepsilon}$ está dada por una función conocida $\mathbf{y} = \boldsymbol{\varphi}(\boldsymbol{\varepsilon})$. En este caso en particular, esa función es

$$y_t = \mathbf{X}_t' \boldsymbol{\beta} + \varepsilon_t$$

Se encuentra que

$$M_{\mathbf{X}_t' \boldsymbol{\beta} + \varepsilon_t}(\theta) = \int_{-\infty}^{\infty} e^{\mathbf{X}_t' \boldsymbol{\beta} + \varepsilon_t} f(\varepsilon) d\varepsilon = e^{(\mathbf{X}_t' \boldsymbol{\beta})\theta} M_{\varepsilon}(\theta)$$

Así, la función generatriz de momentos de $\mathbf{y}$ se determina en términos de la función generatriz de momentos de $\boldsymbol{\varepsilon}$, y el problema de los momentos de $\mathbf{y}$ queda resuelto.

En el caso que se está analizando, afortunadamente, se resuelve en forma sencilla, ya que

$$\mathbf{E}(\mathbf{y}) = \mathbf{E}(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}(\boldsymbol{\varepsilon}) = \mathbf{X}\boldsymbol{\beta}$$

$$\mathbf{V}(\mathbf{y}) = \mathbf{E}[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'] = \mathbf{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}$$

Que es la fórmula utilizada en la función de verosimilitud.

Por otra parte, ante un cambio de variable $\mathbf{y} = \varphi(\boldsymbol{\varepsilon})$ donde la función de densidad de $\boldsymbol{\varepsilon}$ es $f(\boldsymbol{\varepsilon})$, la función de densidad $\mathbf{y}$ se calcula como:

$$g(y) = f(\varepsilon) |J(y)| = f(\varepsilon) \left| \frac{\partial \varepsilon_t}{\partial y_t} \right| = f(y_t - \mathbf{X}_t' \boldsymbol{\beta}) \left| \frac{\partial (y_t - \mathbf{X}_t' \boldsymbol{\beta})}{\partial y_t} \right|$$

Donde $J(Y)$ es el Jacobino de la transformación, determinante definido sobre la matriz de derivadas parciales de $\boldsymbol{\varepsilon}$ respecto a $\mathbf{y}$.

En este caso $J(Y) = 1$ y por lo tanto $g(Y) = f(\boldsymbol{\varepsilon})$; lo que implica que la variable endógena Y está independiente e idénticamente distribuida como una normal multivariada, de la siguiente forma:

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}; \sigma_\varepsilon^2 \mathbf{I}_T)$$

Para maximizar la función de verosimilitud con respecto a $\boldsymbol{\beta}$, es necesario maximizar el exponente o minimizar la suma de cuadrados. Tomando logaritmos se obtiene el logaritmo de la función de verosimilitud

$$\ln L = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Aplicando las condiciones, de primer orden, de máximo respecto a los parámetros desconocidos, se tiene

$$\frac{\partial \ln L}{\partial \boldsymbol{\beta}} = \frac{1}{2\sigma_\varepsilon^2} \mathbf{X}' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

$$\frac{\partial \ln L}{\partial \sigma_\varepsilon^2} = \frac{-T}{2\sigma_\varepsilon^2} + \frac{1}{2\sigma_\varepsilon^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0$$

Resolviendo el sistema y remplazando los parámetros por sus estimadores, se obtiene

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

$$\hat{\sigma}_{MV}^2 = \frac{\mathbf{e}' \mathbf{e}}{T} \neq S^2$$

Condiciones de segundo orden

Para ver si se trata de un máximo, se aplican las condiciones de segundo orden,

$$\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} = -\frac{\mathbf{X}' \mathbf{X}}{\sigma_{\varepsilon}^2} \quad \text{con} \quad -E \left(\frac{\partial^2 \ln L}{\partial \beta \partial \beta'} \right) = \frac{\mathbf{X}' \mathbf{X}}{\sigma_{\varepsilon}^2}$$

$$\frac{\partial^2 \ln L}{\partial \beta \partial \sigma_{\varepsilon}^2} = -\frac{\mathbf{X}' \boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^4} \quad \text{con} \quad -E \left(\frac{\partial^2 \ln L}{\partial \beta \partial \sigma_{\varepsilon}^2} \right) = 0$$

$$\frac{\partial^2 \ln L}{\partial (\sigma_{\varepsilon}^2)^2} = \frac{T}{2\sigma_{\varepsilon}^4} - \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^6} \quad \text{con} \quad -E \left(\frac{\partial^2 \ln L}{\partial (\sigma_{\varepsilon}^2)^2} \right) = \frac{T}{2\sigma_{\varepsilon}^4}; \quad \text{ya que} \quad E(\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}) = T\sigma_{\varepsilon}^2$$

Se puede demostrar que la matriz de derivadas segundas

$$\begin{bmatrix} \frac{\partial^2 \ln L}{\partial \beta \partial \beta' \sigma_{\varepsilon}^2} & \frac{\partial^2 \ln L}{\partial \beta \partial \sigma_{\varepsilon}^2} \\ \frac{\partial^2 \ln L}{\partial \beta \partial \sigma_{\varepsilon}^2} & \frac{\partial^2 \ln L}{\partial (\sigma_{\varepsilon}^2)^2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{\sigma_{\varepsilon}^2} (\mathbf{X}' \mathbf{X}) & -\frac{\mathbf{X}' \boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^4} \\ -\frac{\mathbf{X}' \boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^4} & \frac{T}{2\sigma_{\varepsilon}^4} - \frac{\boldsymbol{\varepsilon}' \boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^6} \end{bmatrix}$$

Constituyen una forma cuadrática definida negativa, condición suficiente para la existencia de un máximo, porque

$$-\frac{1}{\sigma_{\varepsilon}^2}(\mathbf{X}'\mathbf{X}) < 0$$

$$\begin{vmatrix} -\frac{1}{\sigma_{\varepsilon}^2}(\mathbf{X}'\mathbf{X}) & -\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^4} \\ -\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^4} & \frac{T}{2\sigma_{\varepsilon}^4} - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^6} \end{vmatrix} = -\frac{(\mathbf{X}'\mathbf{X})}{\sigma_{\varepsilon}^2} \left[\frac{T}{2\sigma_{\varepsilon}^4} - \frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^6} \right] - \left(-\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sigma_{\varepsilon}^4} \right)^2 > 0$$

Por otra parte, la matriz de información es

$$\mathbf{I} \begin{pmatrix} \boldsymbol{\beta} \\ \sigma_{\varepsilon}^2 \end{pmatrix} = \begin{bmatrix} \frac{1}{\sigma_{\varepsilon}^2}(\mathbf{X}'\mathbf{X}) & \mathbf{0} \\ \mathbf{0} & \frac{T}{2\sigma_{\varepsilon}^4} \end{bmatrix}$$

Y su inversa

$$\mathbf{I}^{-1} \begin{pmatrix} \boldsymbol{\beta} \\ \sigma_{\varepsilon}^2 \end{pmatrix} = \begin{bmatrix} \sigma_{\varepsilon}^2(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\sigma_{\varepsilon}^4}{T} \end{bmatrix}$$

Los términos que están fuera de la diagonal principal son iguales a cero e indican que $\boldsymbol{\beta}$ y σ_{ε}^2 se distribuyen independientemente.

Cabe aclarar que el estimador de máxima verosimilitud tiene varianza sesgada, pero goza de todas las propiedades asintóticas deseables. Es consistente, posee normalidad y eficiencia asintótica,

es invariante y su gradiente tiene media nula y varianza igual a la cota de Cramer – Rao para estimaciones eficientes.

La cota de Cramer – Rao se obtuvo al hacer la inversa de la matriz de información

$$I^{-1} \begin{pmatrix} \beta \\ \sigma_{\epsilon}^2 \end{pmatrix} = \begin{bmatrix} \sigma_{\epsilon}^2 (X'X)^{-1} & 0 \\ 0 & \frac{2\sigma_{\epsilon}^4}{T} \end{bmatrix}$$

Ningún otro estimador con normalidad y consistencia asintótica tiene una matriz de varianzas y covarianzas menor que esta.

Se dijo que los estimadores máximos verosímiles son también invariantes. Esto significa que el estimador máximo verosímil de cualquier función continua de β es esta función del estimador máximo verosímil. Es decir, mientras que con el teorema de Gauss – Markov se afirma que el estimador lineal insesgado más eficiente de β es $\hat{\beta}_{MCO}$, ahora se tiene un resultado asintóticamente más significativo; ya que, el estimador más eficiente de $g(\beta)$, donde $g(\beta)$ es cualquier conjunto de funciones continuas, es $g(\hat{\beta}_{MV})$.

Análisis del sesgo

A pesar de tener una varianza estimada sesgada, ésta solo difiere de S^2 por el factor $-\frac{k}{T}$, ya que el estimador de máxima verosimilitud está sesgado hacia cero. Esto se observa a continuación

$$E(\hat{\sigma}_{MV}^2) = \frac{(T-k)}{T} \sigma^2 = \left(1 - \frac{k}{T}\right) \sigma^2 < \sigma^2$$

Pero el factor $-\frac{k}{T}$ desaparece en muestras grandes.

Asimismo, es posible verificar la equivalencia entre ambos estimadores, aunque sea asintóticamente. A partir de lo analizado, se sabe que es posible, teniendo en cuenta la inversa de la matriz de información y la $E(S^2) = \sigma^2$, escribir

$$T^{1/2}(\hat{\sigma}_{MV}^2 - \sigma^2) \xrightarrow{d} N(0_k, 2\sigma^4)$$

Donde $T^{1/2}(\hat{\sigma}_{MV}^2 - \sigma^2)$ es una variable que representa convenientemente la diferencia de medias de los dos estimadores de la varianza y que, tomando esperanza matemática y varianza sobre la misma, cuando $T \rightarrow \infty$, la media tiende a cero y la varianza a $2\sigma^4$. Tomando esperanza matemática,

$$\begin{aligned} E[T^{1/2}(\hat{\sigma}_{MV}^2 - \sigma^2)] &= T^{1/2}E(\hat{\sigma}_{MV}^2 - \sigma^2) = T^{1/2}[E(\hat{\sigma}_{MV}^2) - E(\sigma^2)] = \\ &= T^{1/2}\left[\left(1 - \frac{k}{T}\right)\sigma^2 - \sigma^2\right] = T^{1/2}\left[-\frac{k\sigma^2}{T}\right] = -\frac{k\sigma^2}{T} \end{aligned}$$

Calculando la varianza de la variable, se tiene

$$\begin{aligned} V[T^{1/2}(\hat{\sigma}_{MV}^2 - \sigma^2)] &= (T^{1/2})^2 V(\hat{\sigma}_{MV}^2 - \sigma^2) = T[V(\hat{\sigma}_{MV}^2) - V(\sigma^2)] = \\ &= T\left[2\frac{\sigma^4}{T}\right] = 2\sigma^4 \end{aligned}$$

Entonces, cuando $T \rightarrow \infty$; $T^{1/2}(\hat{\sigma}_{MV}^2 - \sigma^2) \rightarrow N(0_k, 2\sigma^4)$

Definiendo ahora,

$$z_T = \left(1 - \frac{k}{T}\right) T^{1/2} (\hat{\sigma}_{MV}^2 - \sigma^2) + \frac{k}{T^{1/2}} \sigma^2,$$

Observación. z_T representa una variable centrada y corregida por el sesgo $\left(1 - \frac{k}{T}\right)$ y donde se ha utilizado la deducción anterior de esperanza matemática para centrarla.

De lo cual se puede deducir que la distribución límite de z_T es,

$$\left(1 - \frac{k}{T}\right) N(0_k, 2\sigma^4) + \frac{k}{T^{1/2}} \sigma^2$$

Pero, $\left(\frac{k}{T}\right)$ y $\frac{k}{T^{1/2}}$ desaparecen a medida que $T \rightarrow \infty$, por lo que la distribución límite de z_T también es $N(0_k, 2\sigma^4)$.

Por otra parte, se puede demostrar que centrando convenientemente la variable S^2 , se obtiene

$$z_T = T^{1/2} (S^2 - \sigma^2) \sim N(0_k, 2\sigma^4)$$

Por lo que la distribución asintótica de S^2 es la misma que la del estimador de máxima verosimilitud.

El valor de la función de verosimilitud

Sustituyendo los valores estimados máximo verosímiles en la función logarítmica y tomando antilogaritmos, se obtiene el máximo de la función de verosimilitud

$$\ln L(\hat{\beta}, \hat{\sigma}^2) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \frac{\mathbf{e}'\mathbf{e}}{T} - \frac{T}{2\mathbf{e}'\mathbf{e}} \mathbf{e}'\mathbf{e}$$

$$\ln L(\hat{\beta}, \hat{\sigma}^2) = -\frac{T}{2} \ln 2\pi - \frac{T}{2} \ln \frac{\mathbf{e}'\mathbf{e}}{T} - \frac{T}{2}$$

$$L(\hat{\beta}, \hat{\sigma}^2) = (2\pi)^{-\frac{T}{2}} \left(\frac{\mathbf{e}'\mathbf{e}}{T} \right)^{-\frac{T}{2}} e^{-\frac{T}{2}}$$

$$L(\hat{\beta}, \hat{\sigma}^2) = (2\pi \cdot e)^{-\frac{T}{2}} \left(\frac{\mathbf{e}'\mathbf{e}}{T} \right)^{-\frac{T}{2}}$$

$$L(\hat{\beta}, \hat{\sigma}^2) = \left(\frac{2\pi \cdot e}{T} \right)^{-\frac{T}{2}} (\mathbf{e}'\mathbf{e})^{-\frac{T}{2}}$$

$$L(\hat{\beta}, \hat{\sigma}^2) = \text{constante} \cdot (\mathbf{e}'\mathbf{e})^{-\frac{T}{2}}$$

Donde la constante no depende de ninguno de los parámetros del modelo. La misma depende de las constantes matemáticas π y e .

CASOS DE ESTUDIO, PREGUNTAS Y PROBLEMAS

Caso 15.1: Correlación de muestras

La siguiente tabla proporciona los valores de las medias y las desviaciones estándar de dos variables X e Y , y la correlación de ellas para cada una de las submuestras. Calcular la correlación entre X e Y para la muestra compuesta obtenida juntando las dos submuestras. ¿Porqué dicha correlación es menor que cualquiera de las correlaciones que pudieran existir en las submuestras?

Muestra	Número de muestras	$\bar{X}$	$\bar{Y}$	s_X	s_Y	r_{XY}
1	600	5	12	2	3	0.6
2	400	7	10	3	4	0.7

Caso 15.2: Estimación de parámetros

Una muestra de 20 observaciones correspondiente al modelo

$$Y = \alpha + \beta X + \mu$$

en el que las perturbaciones μ se hallan distribuidas normal e independientemente con media cero y varianza constante, ofrece los siguientes datos:

$$\sum Y = 21.9 \quad \sum (Y - \hat{Y})^2 = 86.9 \quad \sum (X - \bar{X})(Y - \bar{Y}) = 86.9$$

$$\sum X = 186.2 \quad \sum (X - \hat{X})^2 = 215.4$$

- Estimar α y β
- Calcular sus errores estándar.

Caso 15.3: Consumo de cerveza y mortalidad infantil

Un investigador se muestra interesado en las dos series siguientes, definidas para el periodo comprendido entre 1935 y 1946.

Año	35	36	37	38	39	40	41	42	43	44	45	46
X, muerte de niños menores de 1 año (000)	60	62	61	55	53	60	63	53	52	48	49	43
Y, consumo de cerveza (barriles)	23	23	25	25	26	26	29	30	30	32	33	31

- Calcular el coeficiente de correlación entre X e Y.
- Ajustar a X (o Y) una tendencia temporal lineal calculando una regresión MCO de X (o Y) sobre el tiempo t . El procedimiento requiere elegir un origen y una unidad de medida para la variable t . Por ejemplo, estableciendo el origen en la mitad de 1935 y tomando como unidad de medida un año, al año 1942 le corresponderá el valor $t=7$, y así sucesivamente para los demás años. Si el origen se sitúa

a finales de 1940 (principios de 1941) y la unidad de medida es 6 meses, entonces al año 1933 le corresponderá el valor $t=-7$. Demostrar que cualquier tendencia calculada mediante $X_t = a + bt$ no queda afectada por la elección del origen y la unidad de medida.

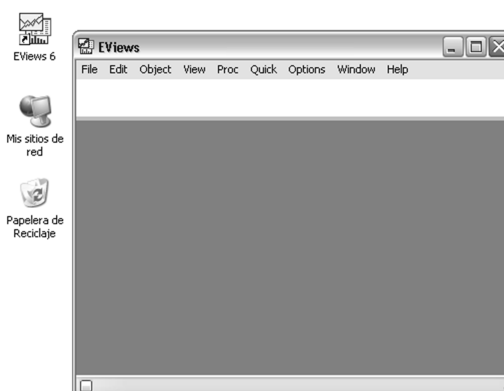
- c) Suponer que $e_{X,t}$ y $e_{Y,t}$ indican los residuos de X e Y respecto a sus valores tendenciales. Calcular los coeficientes de correlación entre $e_{X,t}$ y $e_{Y,t}$. Comparar dicho valor con el obtenido en el apartado a) y comentar la justificación de tales diferencias.

Caso 15.4: Primeros pasos en Eviews

Eviews 6 es un software de la empresa *Quantitative Micro Software* (<http://www.eviews.com/>) orientado al análisis econométrico. Conjuntamente con STATA, es uno de los paquetes disponibles en la actualidad más usados y más. A diferencia de otros programas conocidos como SAS o SPSS -que se orientan al análisis estadístico general- Eviews se especializa en econometría, tanto de series de tiempo como de corte transversal y datos de panel.

Posee la ventaja de contar con una *interface* gráfica, de modo que resulta mucho más intuitivo que otras alternativas; pero también cuenta con el potencial de un entorno programable para usuarios avanzados, como son S o R (versión *freeware* de S). Una alternativa *freeware* a Eviews es "gretl", un paquete econométrico con *interface* usuario gráfica (http://gretl.sourceforge.net/gretl_espanol.html).

En el escritorio de la PC, seguramente se encuentra un icono como el de la figura, y tras abrirlo, una ventana con fondo grisáceo sin más detalles. Se comenzará utilizando datos de la Tabla 15.1 del Ejemplo 15.5.a.

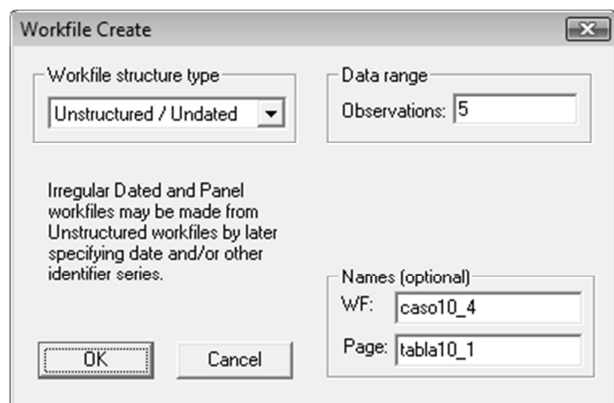


Icono y ventana inicial de Eviews 6

Creando Workfile.

El archivo base con el cual trabaja este programa es el *workfile* (fichero de trabajo) que se crea desde el menú *File > New > workfile...* (archivo>nuevo>fichero de trabajo). En primer lugar se abre la ventana *Workfile create* donde se debe especificar la estructura del archivo.

Las opciones disponibles son: (1) *Unstructured / Undated* (Desestructurado / No Fechado) que se utiliza en caso de datos que no se corresponden con observaciones en el tiempo regulares –los datos del ejemplo, existentes en tabla 15.1, no se corresponden



Creación de un nuevo archivo

con algún período de tiempo-; (2) *Dated - Regular Frequency* (Fechado – Frecuencia Regular) para tabla de datos donde las unidades de observación se corresponden con

unidades regulares de tiempo -como años, trimestres, meses, etc., se debe especificar la frecuencia, *start date* (fecha inicial) y *end date* (Fecha final)-; y (3) *Balanced Panel* (Panel Balanceado) cuando para cada individuo observado se dispone de series de tiempo de igual longitud, tal que deben especificarse frecuencia, fecha inicial, fecha final y número de *cross-section* (secciones cruzadas).

Puesto que en la Tabla 15.1 $i = (1, 2, \dots, 5)$, en *Data Range* (Rango de datos) se indica 5 observaciones. También es posible, aunque no es necesario, indicar el nombre de archivo en *WF* y el nombre de hoja (como en Excel) en *Page*.

Cargando datos

Para cargar los datos en el archivo creado existen dos maneras diferentes: (1) importando desde una aplicación externa, como puede ser Microsoft® Excel, o (2) tipeando directamente en Eviews.

	A	B	C	D
1	t	Y	X2	X3
2	1	7	4	2
3	2	10	6	4
4	3	5	2	1
5	4	11	5	10
6	5	20	16	3

Tabla de Datos en Excel

El primero de los métodos consiste en utilizar una planilla de cálculo para cargar los datos que luego se guardan con alguna de las siguientes extensiones: **.xls* (Excel 97-2003), **.wks* (Lotus), otros

archivos de texto ASCII como **.txt* y **.csv*. Una vez confeccionada la Tabla 15.1 y guardada con el nombre *tabla15_1.xls*, hay que

asegurarse de cerrarla y de que ningún programa la esté utilizando.

En Eviews, desde el menú *File > Import > Read Text-Lotus-Excel* (Archivo > Importar > Leer Texto-Lotus-Excel) se abre el cuadro de diálogo para encontrar la ubicación del archivo con la tabla.

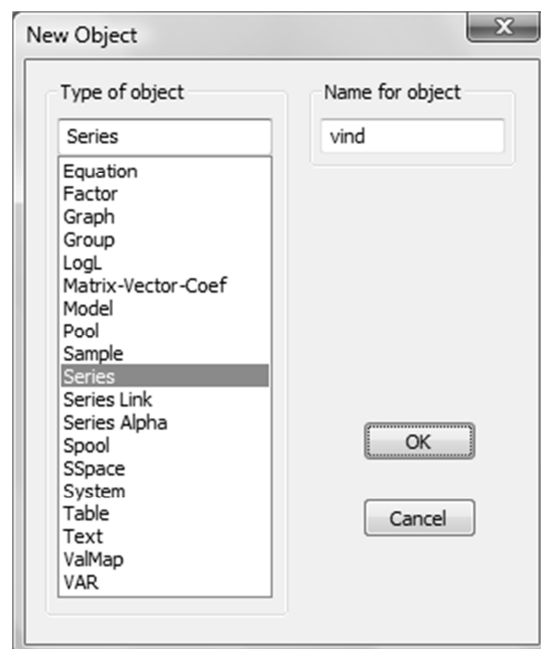
Primero, hay que indicarle al programa como están ubicados los datos en el archivo. En el caso de la tabla 15.1 las filas son las observaciones, por lo que se marca la opción *By Observation – series in columns*. En *Upper-left data cell*, hay que indicar la celda a partir de la cual comienzan los datos propiamente dichos, o sea la celda *B2*, pues la columna *A* contiene rótulos de observaciones, y la Fila *1* los rótulos para las variables. En *names for series or number if named in file* hay que detallar el nombre de las series de la tabla o bien, si deseamos que el programa importe los nombres originales, indicar cuantas variables contiene la tabla. Para este ejercicio se puede escribir *vdep vind1 vind2* para renombrar a *Y*, *X1*, *X2* respectivamente.

obs	VIND1	VIND2	VDEP
1	2.000000	1.000000	4.000000
2	3.000000	3.000000	7.000000
3	1.000000	5.000000	3.000000
4	5.000000	1.000000	9.000000
5	9.000000	2.000000	17.000000

Si se han realizado bien todos los pasos, el workfile debería contener cinco observaciones y tres variables: *vind1*, *vind2*, *vdep*; además de reservar el espacio para el vector de coeficientes estimados (*c*) y la serie de los residuos (*resid*). Una vez importados

los datos, es posible verificarlos seleccionando los íconos de las variables, *cliqueando* con el botón derecho del mouse y eligiendo *Open > as group*.

El segundo método para incluir datos consiste en generar series mediante el menú *Object > New Object*. Se debe especificar *Type of Object: Series* y un nombre. Una vez generados los objetos se abren en grupo o individualmente, presionando *Edit +/-* es posible tipear los datos como si fuera un planilla de cálculo normal.



Cuadro de Dialogo Crear Series

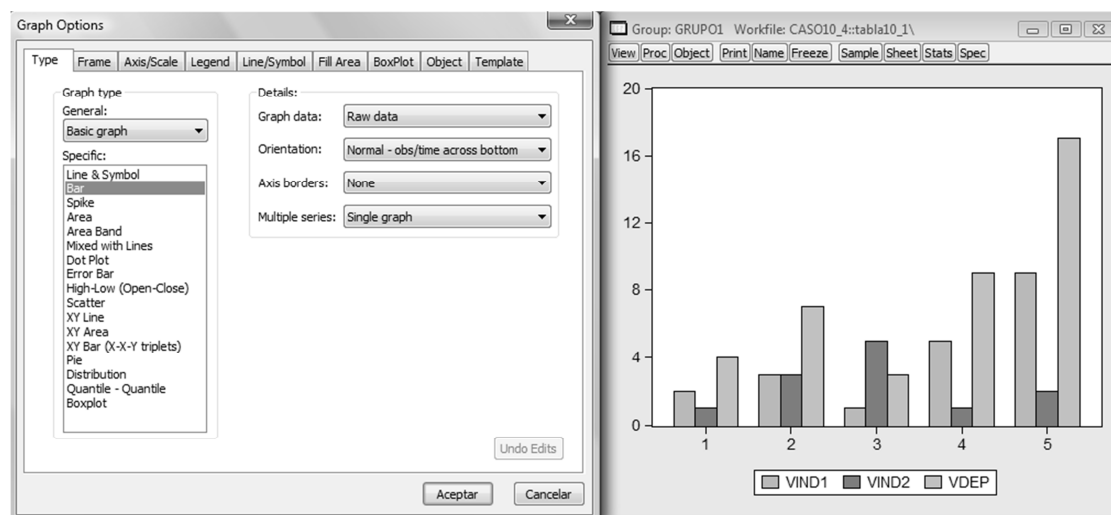
Trabajando con los Datos

Los Grupos abiertos pueden guardarse con un nombre para encontrarlos fácilmente después. Basta con seleccionar el botón *Name* y escribir el nombre deseado. Otras herramientas importantes con las que pueden trabajar dentro de la ventana del grupo se encuentran en el menú *View*. Así en *Group Members* obtienen el listado de variables que observan:

Edit series expressions below this line -- 'UpdateGroup' applies edits to Group.

VIND1
VIND2
VDEP

La opción *Spreadsheet* les permite volver a la planilla con los datos. La opción *Graph* les permite graficar los datos en un gran número de formas diferentes; pueden elegir el tipo de gráfico en la primera pestaña del cuadro de diálogo y pueden cambiar el aspecto del gráfico en las pestañas restantes. Como ejemplo, dado que los datos no tienen estructura temporal puede preferirse verlos en forma de barras y no como curvas; para esto se elige *General: Basic Graph*, luego *Specific: Bar* y, para visualizar las tres series en un mismo gráfico, luego *Multiple Series: Single Graph*.



Graficando Series

Otra herramienta importante -que se debe utilizar cuando se comienza a trabajar con los datos- son las estadísticas descriptivas. Haciendo *View > Descriptive Stats > Common Sample* (muestra común) el programa genera una tabla con las estadísticas para cada variable: Media (mean), mediana (median), máximo (maximum), mínimo (minimum), desviación estándar (std. Dev.), asimetría (skewness), el estadístico Kurtosis, Jarque – Bera con su probabilidad, la suma y la suma de desvíos cuadráticos (sum sq. Dev).

También pueden efectuar Análisis de Covarianza (*Covariance Analysis*), el que les permite visualizar la matriz de covarianzas, la matriz de correlación y asociar a estas la matriz de pruebas t para hipótesis de covarianza nula o independencia. Otro tipo de pruebas de hipótesis son los test de igualdad (*test equality*) para medias, medianas y varianzas a los que acceden a través del menú View.

Análisis de Regresión

Para realizar una explicación del comportamiento de la variable dependiente se especifica el siguiente modelo

$$Vdep_i = \beta_1 + \beta_2 Vind1_i + \beta_3 Vind2_i + \varepsilon_i \quad \forall 1, 2, \dots, 5$$

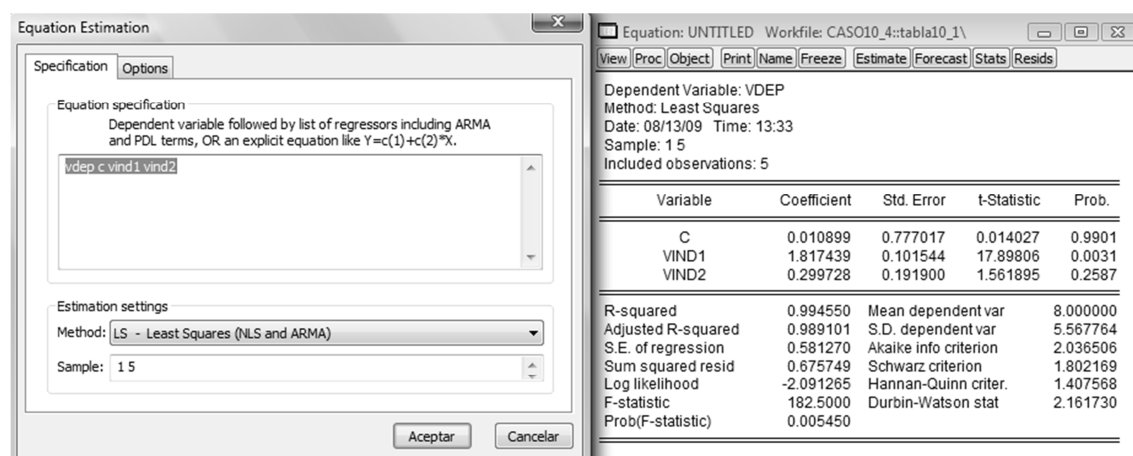
El valor de los parámetros de la Regresión se obtiene desde el menú *Quick > Estimate Equation*, en el cuadro de diálogo se especifica la estimación escribiendo

$$vdep \ c \ vind1 \ vind2$$

El término *c* indica que se debe calcular la constante de la regresión; si hubiera más variables explicativas, se consignan a continuación. En *Estimation settings* se elige *Last Square* (Mínimos Cuadrados) y en *Sample* (muestra) se escribe *1 5*, es decir desde la observación 1° a la 5°. Una vez que se acepta esta configuración se obtiene la "Salida" de la Estimación (*Estimation Output*).

El primer grupo de elementos de la salida indica la variable a explicar, el método empleado, la muestra considerada y la cantidad total de observaciones consideradas; este último dato podría se

menor que el tamaño de muestra pues podrían faltar datos o establecerse una especificación del modelo a estimar que imposibilitara utilizar todas las observaciones.



Cuadro de Dialogo Estimar ecuación y Salida de la Regresión.

El segundo grupo contiene la estimación de los coeficientes, sus errores estándar y la prueba t con la significatividad correspondiente.

El tercer grupo de información contiene estadísticos útiles para evaluar la bondad del ajuste de la regresión, la significatividad conjunta y la calidad de la estimación en cuanto al cumplimiento de los supuestos básicos del modelo lineal general.

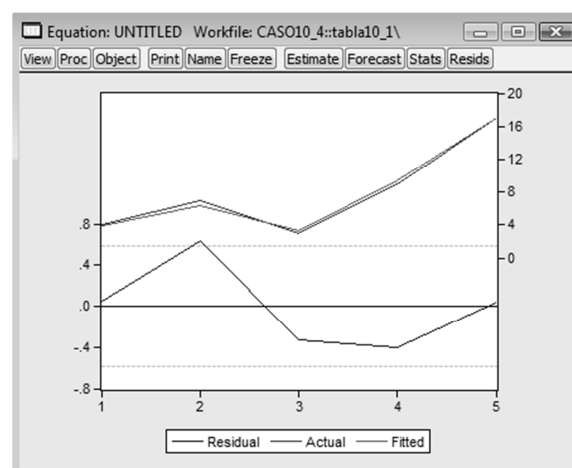


Gráfico de la Variable, su estimación y los errores

Finalmente, desde esta misma ventana de estimación pueden plotearse gráficos para la variable dependiente, los valores estimados y los errores de estimación. Para ello se selecciona el menú *View > Actual, Fitted, Residual > Actual, Fitted, Residual graph* (gráfico real, estimado, y de residuos). Desde el mismo menú *View* se accede a los test y pruebas de hipótesis sobre el modelo estimado que se estudiará a lo largo de la materia.

Actividades Propuestas

- a) Realice todos los pasos comentados anteriormente para familiarizarse con el manejo del software.
- b) Compare la información de la salida de la estimación, con los resultados obtenidos a lo largo del Capítulo 15 en relación al ejemplo 15.5.
- c) Interprete, con los conocimientos ya aprendidos y los que recuerda de Inferencia Estadística, el significado de la información contenida en la salida.
- d) Localice en la Salida el estadístico

$$SCR = \sum_{t=1}^T e_t^2 = \mathbf{e}'\mathbf{e}$$

- e) Repitiendo los pasos explicados en este caso, seleccione un modelo económico con el que haya trabajado teóricamente, especifique el modelo econométrico, busque los datos y estime un modelo de regresión lineal.
-
-

Bibliografía

- **Chiang, Alpha.** *Métodos Fundamentales De Economía Matemática*. México: McGraw Hill, 1987.
- **Gujarati, Damodar.** *Econometría*. México: Mc.Graw Hill, 2004.
- **Johnston, J. y Dinardo, J.** *Métodos De Econometría*. Barcelona: Editorial Vicens Vives, 2001.
- **Pulido San Román, Antonio.** *Modelos Econométricos*. Madrid: Editorial Pirámide, 1993.