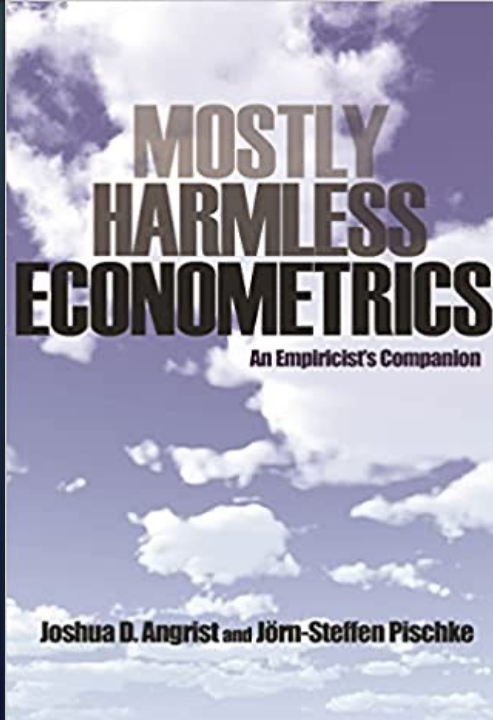
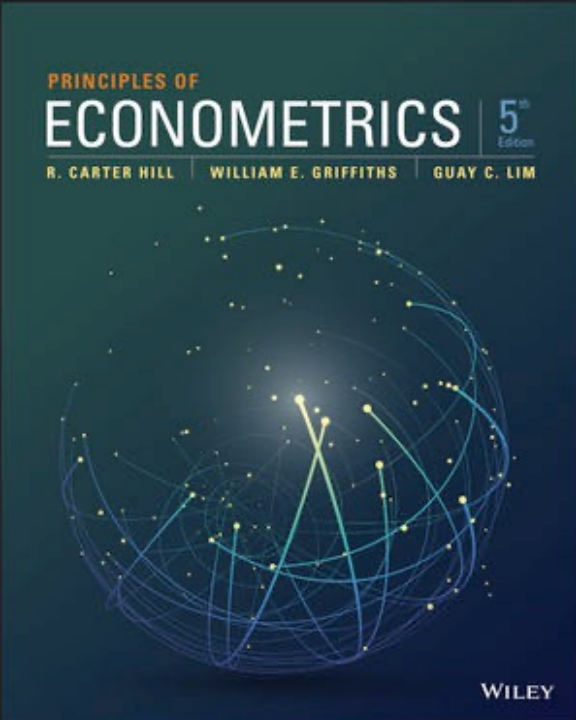
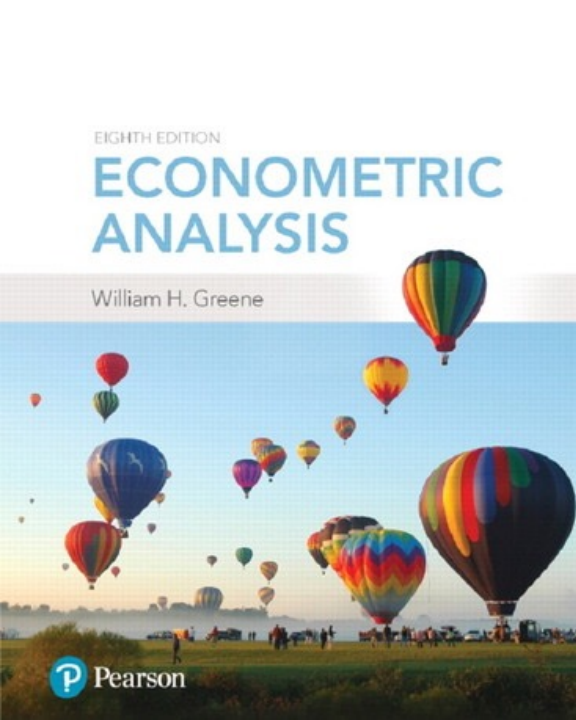




PUCP



DEPARTAMENTO ECONOMÍA
FUNDAMENTOS DE ECONOMETRÍA
1ECO11 – HORARIO 0723

Sesión 5

Regresión Múltiple

Docente: Juan Palomino



ÍNDICE

1

Modelo con k variables

2

Álgebra de MCO

3

Propiedades Matemáticas de MCO

4

Propiedades Estadísticas Finitas de MCO

5

El Modelo en Desviaciones respecto a la Media

6

Omisión de Variables Relevantes

7

Inclusión de Variables Irrelevantes

8

Teorema de Frisch-Waugh

9

Descomposición de la Suma de Cuadrados

1. Modelo con k variables

Modelo con k variables

La extensión del modelo bivariado es la siguiente:

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ik} + \varepsilon_i \quad i = 1, \dots, n$$

En términos matriciales:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & X_{31} & \cdots & X_{k1} \\ X_{12} & X_{22} & X_{32} & \cdots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & X_{3n} & \cdots & X_{kn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

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

Supuesto 1: Linealidad

La relación entre la variable dependiente y las variables independientes es lineal. El modelo poblacional es:

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K X_{iK} + \varepsilon_i \quad i = 1, \dots, n$$

Donde β 's son parámetros desconocidos a ser estimados, y ε_i son términos de errores observados.

Nota: La propiedad de linealidad es una propiedad de los parámetros, no de las variables.

Supuesto 2: Exogeneidad Estricta

El valor esperado de la perturbación aleatoria debe ser cero para cualquier observación:

$$E[\varepsilon|X] = \begin{bmatrix} E[\varepsilon_1|X] \\ E[\varepsilon_2|X] \\ \vdots \\ E[\varepsilon_n|X] \end{bmatrix} = 0 \quad i = 1, 2, \dots, n$$

Para cualquier valor de X_1, X_2, \dots, X_K en la población, el promedio no observable es igual a cero.

Supuestos del Modelo Clásico: Exogeneidad Estricta

Consecuencias de Exogeneidad Estricta:

Definición: Ley de Esperanzas Iteradas

Si $E|Y| < \infty$, es decir, la esperanza poblacional de Y existe, entonces para cualquier vector X aleatorio:

$$E_X[E(Y|X)] = E(Y)$$

1. Media Incondicional

La media incondicional del término de error es cero:

$$E(\varepsilon) = 0$$

Demostración. Por la Ley de Esperanzas Iteradas (LIE)

$$E(\varepsilon) = E[E(\varepsilon|X)] = 0$$

Consecuencias de Exogeneidad Estricta:

2. Regresores son ortogonales al término de error

Bajo exogeneidad estricta, los regresores son ortogonales al término de error para todas las observaciones, es decir, no comparten información.

$$E(X_{jk}\varepsilon_i) = 0 \quad i, j = 1, 2, \dots, n \quad k = 1, \dots, K$$

O

$$E(X \cdot \varepsilon_i) = \begin{pmatrix} E(X_{j1}\varepsilon_i) \\ E(X_{j2}\varepsilon_i) \\ \vdots \\ E(X_{jK}\varepsilon_i) \end{pmatrix} = \underbrace{0}_{K \times 1} \quad \forall i, j$$

Supuestos del Modelo Clásico: Exogeneidad Estricta

Consecuencias de Exogeneidad Estricta:

3. Condición de cero correlación

El regresor no está correlacionado con el término de error:

$$\text{Cov}(\varepsilon, X) = 0$$

4. Media condicional de y

La media condicional de la variable dependiente es una función lineal del regresor.

Bajo el supuesto de linealidad y exogeneidad estricta:

$$E(y|X) = E(X\beta + \varepsilon|X) = X\beta + E(\varepsilon|X)$$

$$E(y|X) = X\beta$$

Consecuencias de Exogeneidad Estricta:

3. Condición de cero correlación

El regresos no está correlacionado con el término de error:

$$\text{Cov}(\varepsilon_i, x_{jk}) = 0$$

4. Media condicional de y_i

La media condicional de la variable dependiente es una función lineal de los regresores.

Bajo el supuesto de linealidad y exogeneidad estricta:

$$E(y_i|X) = x_i'\beta \quad i = 1, 2, \dots, n$$

Supuesto 3: Rango Completo

Asume que ninguna de las variables en el modelo es una combinación lineal exacta de las otras.

X es una matriz $n \times K$ con rango K

Significa que X tiene rango de columna completo.

Nota: El rango de una matriz es igual al número de columnas linealmente independientes de la matriz. También se le conoce como **condición de identificación**.

Supuesto 4: Perturbaciones Esféricas

❖ Homocedasticidad

Asume que todas las unidades tienen el mismo error de varianza

$$E[\varepsilon_i^2 | X] = \sigma_0^2 > 0 \quad i = 1, 2, \dots, n$$

❖ No Correlación

Asume también que no hay correlación entre las observaciones

$$E[\varepsilon_i \varepsilon_j | X] = 0 \quad i, j = 1, 2, \dots, n \quad i \neq j$$

Perturbaciones Esféricas

Este supuesto señala que la varianza de todos los términos de errores es constante.

Para ver esto:

$$Var(\varepsilon_i|X) = E(\varepsilon_i^2|X) - [E(\varepsilon_i|X)]^2 \quad \text{(Por varianza condicional)}$$

$$Var(\varepsilon_i|X) = E(\varepsilon_i^2|X) \quad \text{(Por exogeneidad estricta)}$$

Similarmente, la covarianza de la distribución conjunta de $(\varepsilon_i \varepsilon_j)$ condicional a X es cero.

$$Cov(\varepsilon_i \varepsilon_j | X) = 0 \quad (i, j = 1, 2, \dots, n \quad i \neq j)$$

$$E(\varepsilon_i \varepsilon_j | X) - E(\varepsilon_i | X)E(\varepsilon_j | X) = 0$$

$$E(\varepsilon_i \varepsilon_j | X) = 0$$

En notación matricial:

$$E(\varepsilon\varepsilon'|X) = \begin{pmatrix} E(\varepsilon_1\varepsilon_1|X) & E(\varepsilon_1\varepsilon_2|X) & \cdots & E(\varepsilon_1\varepsilon_n|X) \\ E(\varepsilon_2\varepsilon_1|X) & E(\varepsilon_2\varepsilon_2|X) & \cdots & E(\varepsilon_2\varepsilon_n|X) \\ \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_n\varepsilon_1|X) & E(\varepsilon_n\varepsilon_2|X) & \cdots & E(\varepsilon_n\varepsilon_n|X) \end{pmatrix}$$

$$E(\varepsilon\varepsilon'|X) = \begin{pmatrix} \sigma_0^2 & 0 & \cdots & 0 \\ 0 & \sigma_0^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_0^2 \end{pmatrix} = \sigma_0^2 I_n$$

Los términos de errores que cumplen los supuestos de homocedasticidad y no autocorrelación son llamados “**perturbaciones o errores esféricas**”.

Supuesto 5: Regresores no estocásticos

Las observaciones de las variables exógenas X son fijas en muestras repetidas.

Asumir que los X son fijos quiere decir que, en repetidas muestras de X , los valores obtenidos X_1, X_2, \dots, X_K van a ser siempre los mismos; es decir, dejan de ser aleatorios.

Bajo este supuesto, ya no es necesario hablar de esperanzas condicionales:

$$E(\varepsilon) = 0$$

$$\text{Var}(\varepsilon) = \sigma^2$$

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0$$

Supuesto 6: Normalidad de los errores

ε distribuye normal con media cero y varianza σ^2 condicional a X :

$$\varepsilon | X \sim N(0, \sigma^2), \quad i = 1, 2, \dots, n$$

2. Álgebra de MCO

La función de regresión muestral en matrices es:

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

Definimos el vector de residuos como:

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

La suma de residuos al cuadrado (SCR):

$$SCR(\hat{\beta}) = \sum_{i=1}^n (y - X\hat{\beta})^2 = \hat{\varepsilon}'\hat{\varepsilon} = (y - X\hat{\beta})'(y - X\hat{\beta})$$

Derivamos las condiciones de primer orden mediante derivadas parciales que sean iguales a cero.

Entonces:

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

$$SCR(\hat{\beta}) = (y' - \hat{\beta}'X') (y - X\hat{\beta}) \quad \text{ya que } (X\hat{\beta})' = \hat{\beta}'X'$$

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

$$SCR(\hat{\beta}) = y'y - 2y'X\hat{\beta} - \hat{\beta}'X'X\hat{\beta})$$

Nota: $\hat{\beta}'X'y = (y'X\hat{\beta})'$ y el término $y'y$ no depende de β y puede ser ignorado en la diferenciación de $SCR(\hat{\beta})$

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

$$-2X'y - 2X'X\hat{\beta} = 0$$

$$2X'X\hat{\beta} = 2X'y$$

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

$K \times 1 \quad K \times K \quad K \times 1$

Las $K \times 1$ ecuaciones:

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

Son llamadas **ecuaciones normales**.

3. Propiedades Matemáticas de MCO

❖ Propiedad

X y ε son ortogonales:

$$X'(y - X\hat{\beta}_{OLS}) = 0 \quad X'\hat{\varepsilon} = 0$$

que es lo mismo que:

$$\frac{1}{n} \sum_{i=1}^n x_i \cdot \hat{\varepsilon}_i = 0 \quad \frac{1}{n} \sum_{i=1}^n x_i \cdot (y - X\hat{\beta}_{OLS}) = 0$$

Dos expresiones para el estimador de MCO

Por el supuesto de no multicolinealidad, la matriz $X'X$ es definida positiva y por lo tanto no singular.

Entonces:

$$(X'X)^{-1}X'y = \left(\frac{X'X}{n}\right)^{-1} \left(\frac{X'y}{n}\right)$$

El estimador de MCO puede ser reescrito como:

$$\hat{\beta}_{OLS} = S_{XX}^{-1}S_{Xy}$$

Donde:

$$S_{XX} = \frac{1}{n}X'X = \frac{1}{n}\sum_{i=1}^n x_i x_i'$$

Promedio muestral de $x_i x_i'$

$$S_{Xy} = \frac{1}{n}X'y = \frac{1}{n}\sum_{i=1}^n x_i y_i$$

Promedio muestral de $x_i y_i$

Matriz de Proyección y Aniquiladora

La matriz proyección P y la matriz aniquiladora M son definidos como:

$$P = X(X'X)^{-1}X'$$

$n \times n$

$$M = I_n - P$$

$n \times n$

Las propiedades de las matrices P y M :

1. Tanto P y M son simétricos e idempotentes.
2. $PX = X$, de ahí el término de matriz de proyección.
3. $MX = 0$, de ahí el término aniquilador.

Ya que $\hat{\varepsilon}_i$ es el vector residual, la suma de los residuos al cuadrado, SCR , es igual a $\hat{\varepsilon}'\hat{\varepsilon}$. Se puede escribir así:

$$SCR = \hat{\varepsilon}'\hat{\varepsilon} = \hat{\varepsilon}'M\hat{\varepsilon}$$

Varianza del término de error

El estimador MCO de σ^2 (la varianza del término de error), denotado por $\hat{\sigma}^2$, es la suma de residuos al cuadrado dividido por $n - K$:

$$\hat{\sigma}^2 = \frac{SCR}{n - K} = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{n - K}$$

Donde $n - K$ son los grados de libertad.

4. Propiedades Estadísticas de MCO

Insesgadez del Estimador de MCO

Estimador Insesgado

Un estimador $\hat{\theta}$ para θ_0 es insesgado si $E(\hat{\theta}) = \theta_0$

Teorema 1: Insesgadez del Estimador MCO

Bajo el supuesto de Linealidad, Exogeneidad Estricta y Rango Completo:

$$E(\hat{\beta}|X) = \beta_0$$

Entonces, $\hat{\beta}$ es un estimador insesgado de β_0

Insesgadez del Estimador de MCO

Ejemplo:

Tenemos el siguiente modelo:

$$bmi_i = \alpha_0 + \beta_0 income_i + \varepsilon_i$$

Donde β_0 es el efecto del ingreso sobre obesidad y es igual a 1.

El siguiente paso es conseguir una muestra de la población y obtener un estimado de β_0 , llamado $\hat{\beta}$.

Imaginen lo siguiente: $B = 50$ muestras de 1000 individuos y para cada muestra estimamos β .

$$\text{1era muestra} \rightarrow \hat{\beta}_1 = 0.9$$

$$\text{2da muestra} \rightarrow \hat{\beta}_1 = 0.94$$

$$\text{3era muestra} \rightarrow \hat{\beta}_1 = 0.7$$

$$\text{50ava muestra} \rightarrow \hat{\beta}_1 = 0.5$$

En promedio de las 50 muestras, conseguir: $(\frac{1}{B}) \sum_{b=1}^B \hat{\beta}_b \approx 1$

Demostración

Tenemos:

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

$$\hat{\beta} = (X'X)^{-1}X'(X\beta_0 + \varepsilon)$$

$$\hat{\beta} = (X'X)^{-1}X'X\beta_0 + (X'X)^{-1}X'\varepsilon$$

$$\hat{\beta} = \beta_0 + (X'X)^{-1}X'\varepsilon$$

$$\hat{\beta} - \beta_0 = A\varepsilon$$

Donde $A = (X'X)^{-1}X'$. Se le conoce como **sampling error**.

Teorema 2: Varianza del estimador MCO

Bajo el supuesto de Linealidad, Exogeneidad Estricta, Rango Completo y Errores Esféricos:

$$\text{Var}(\hat{\beta}_{OLS}) = \sigma_0^2 (X'X)^{-1}$$

Entonces, $\hat{\beta}$ es un estimador insesgado de β_0

Consecuencias de relajar el supuesto: Varianza

Demostración: Varianza Homocedástica

Tenemos:

$$Var(\hat{\beta}|X) = E(\hat{\beta} - E[\hat{\beta}|X])^2 \quad \text{(Por varianza condicional)}$$

$$Var(\hat{\beta}|X) = E(\hat{\beta} - \beta_0|X)^2 \quad \text{(Por insesgadez)}$$

$$Var(\hat{\beta}|X) = E(A\varepsilon|X)^2 \quad \text{(Por sampling error)}$$

$$Var(\hat{\beta}|X) = E(A\varepsilon\varepsilon'A'|X) \quad \text{(Forma cuadrática)}$$

$$Var(\hat{\beta}|X) = AE(\varepsilon\varepsilon'|X)A' \quad \text{(A es una función de X)}$$

$$Var(\hat{\beta}|X) = A(\sigma_0^2 I_n)A' \quad \text{(Por homocedasticidad)}$$

$$Var(\hat{\beta}|X) = \sigma_0^2 AA'$$

$$Var(\hat{\beta}|X) = \sigma_0^2 (X'X)^{-1}$$

Teorema 3: Gauss-Markov

Bajo todos los supuestos, el estimador MCO es eficiente en la clase de estimadores lineales insesgados. Es decir, para cualquier estimador insesgado $\tilde{\beta}$ que es lineal en y ,

$$\text{Var}(\tilde{\beta}) \geq \text{Var}(\hat{\beta}_{OLS}|X)$$

Este teorema señala que el estimador MCO es eficiente en el sentido que su matriz de varianza-covarianza $\text{Var}(\hat{\beta}_{OLS}|X)$ es el más pequeño entre todos los estimadores insesgados. Por esta razón, al estimador MCO se le conoce como Mejor Estimador Lineal Insesgado (MELI) [Best Linear Unbiased Estimator (BLUE)].

Teorema de Gauss Markov

Demostración:

Ya que $\tilde{\beta}$ es lineal en y , se puede escribir como $\tilde{\beta} = Cy$ para alguna matriz C de dimensión $n \times K$, el cual está posiblemente en función de X .

Sea $D = C - A$ o $C = D + A$, donde $A = (X'X)^{-1}X'$. Entonces:

$$\tilde{\beta} = (D + A)y$$

$$\tilde{\beta} = Dy + Ay$$

$$\tilde{\beta} = D(X\beta + \varepsilon) + \hat{\beta}$$

$$\tilde{\beta} = DX\beta + D\varepsilon + \hat{\beta}$$

Teorema de Gauss Markov

Demostración:

Tomando esperanza condicional ambos lados, se obtiene:

$$E(\tilde{\beta}|X) = DX\beta + E(D\varepsilon|X) + E(\hat{\beta}|X)$$

Ya que tanto $\hat{\beta}$ y $\tilde{\beta}$ son insesgados y que:

$$E(D\varepsilon|X) = DE(\varepsilon|X) = 0$$

Resulta que $DX\beta = 0$. Para que esto sea cierto, para cualquier β_0 es necesario que $DX = 0$. Por lo tanto, $\tilde{\beta} = D\varepsilon + \hat{\beta}$ y

$$\begin{aligned}\tilde{\beta} - \beta_0 &= D\varepsilon + (\tilde{\beta} - \beta_0) \\ &= (D + A)\varepsilon\end{aligned}$$

Teorema de Gauss Markov

Demostración:

Por tanto:

$$\text{Var}(\tilde{\beta}|X) = \text{Var}(\tilde{\beta} - \beta_0|X)$$

$$\text{Var}(\tilde{\beta}|X) = \text{Var}[(D + A)\varepsilon|X]$$

$$\text{Var}(\tilde{\beta}|X) = (D + A)\text{Var}(\varepsilon|X)(D + A)'$$

$$\text{Var}(\tilde{\beta}|X) = \sigma^2(D + A)(D' + A')$$

$$\text{Var}(\tilde{\beta}|X) = \sigma^2(DD' + AD' + DA' + AA')$$

Pero $DA' = DX(X'X)^{-1} = 0$ ya que $DX = 0$. También, $AA' = (X'X)^{-1}$. Por tanto:

$$\text{Var}(\tilde{\beta}|X) = \sigma^2[DD' + (X'X)^{-1}]$$

$$\text{Var}(\tilde{\beta}|X) \geq \sigma^2(X'X)^{-1} \text{ ya que } DD' \text{ es semidefinida positiva}$$

$$\text{Var}(\tilde{\beta}|X) \geq \text{Var}(\hat{\beta}|X)$$

Teorema 4: Covarianza entre $\hat{\beta}$ y $\hat{\varepsilon}$

Bajo todos los supuestos de Linealidad, Exogeneidad Estricta, Rango Completo, y Errores Esféricos:

$$\text{Cov}(\hat{\beta}, \hat{\varepsilon} | X) = 0$$

Donde $\hat{\varepsilon} \equiv y - X\hat{\beta}$

Covarianza entre $\hat{\beta}$ y $\hat{\varepsilon}$

Demostración

$$\begin{aligned} \text{Cov}(\hat{\beta}, \hat{\varepsilon}|X) &= E\{[(\hat{\beta} - E(\hat{\beta}|X))][\hat{\varepsilon} - E(\hat{\varepsilon}|X)]'|X\} \\ &= E\{[(\hat{\beta} - \beta_0|X)][M\varepsilon - E(M\varepsilon|X)]'|X\} \\ &= E[A\varepsilon\varepsilon'M|X] \\ &= AE[\varepsilon\varepsilon'|X]M' \\ &= AE[\sigma_0^2 I_n]M \\ &= \sigma_0^2 (X'X)^{-1} X'M \\ &= \sigma_0^2 (X'X)^{-1} \underbrace{(MX)'}_0 \\ &= 0 \end{aligned}$$

Por insesgadez y $M\varepsilon = \hat{\varepsilon}$

Por sampling error: $\hat{\beta} - \beta_0 = A\varepsilon$

Ya que M y A son funciones de X

Teorema 5: Insesgadez de $\hat{\sigma}^2$

Bajo todos los supuestos de Linealidad, Exogeneidad Estricta, Rango Completo, y Errores Esféricos:

$$E(\hat{\sigma}^2|X) = \sigma_0^2$$

Siempre que $n > K$, de modo que $\hat{\sigma}^2$ está bien definido.

Demostración

Ya que $\hat{\sigma}^2 = \frac{\hat{\varepsilon}'\hat{\varepsilon}}{(n-K)}$, la prueba equivale a mostrar que $E(\hat{\varepsilon}'\hat{\varepsilon}|X) = (n-K)\sigma_0^2$. Sabemos que $\hat{\varepsilon}'\hat{\varepsilon} = \varepsilon M \varepsilon$, donde M es la matriz aniquiladora. La demostración consiste en probar dos propiedades: (1) $E(\varepsilon M \varepsilon|X) = \sigma_0^2 \cdot \text{traza}(M)$, y (2) $\text{traza}(M) = n - K$.

1. Probar que $E(\varepsilon M \varepsilon|X) = \sigma_0^2 \cdot \text{traza}(M)$: ya que $\varepsilon' M \varepsilon = \sum_{i=1}^n \sum_{j=1}^n m_{ij} \varepsilon_i \varepsilon_j$, tenemos:

$$E(\varepsilon M \varepsilon|X) = \sum_{i=1}^n \sum_{j=1}^n m_{ij} E(\varepsilon_i \varepsilon_j|X)$$

Ya que m_{ij} 's son funciones de X

$$E(\varepsilon M \varepsilon|X) = \sum_{j=1}^n m_{ii} \sigma^2$$

Ya que $E(\varepsilon_i \varepsilon_j|X) = 0$ para $i \neq j$ por supuesto 4

$$E(\varepsilon M \varepsilon|X) = \sigma^2 \sum_{j=1}^n m_{ii} = \sigma^2 \cdot \text{traza}(M)$$

Demostración

2. Probar que $\text{traza}(M) = n - K$

$$\text{traza}(M) = \text{traza}(I_n - P)$$

Ya que $M = I_n - P$

$$\text{traza}(M) = \text{traza}(I_n) - \text{traza}(P)$$

El operador de la traza es lineal

$$\text{traza}(M) = n - \text{traza}(P)$$

y

$$\text{traza}(P) = \text{traza}[X(X'X)^{-1}X']$$

Ya que $P = X(X'X)^{-1}X'$

$$\text{traza}(P) = \text{traza}[(X'X)^{-1}X'X]$$

Ya que $\text{traza}(AB) = \text{traza}(BA)$

$$\text{traza}(P) = \text{traza}[I_K]$$

$$\text{traza}(P) = K$$

Por lo tanto, $\text{traza}(M) = n - K$

Teorema 6: Varianza de $\hat{\sigma}^2$

La varianza del estimador $\hat{\sigma}^2$ es:

$$\text{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{n - k}$$

Un resultado que involucra al estimador $\hat{\sigma}^2$ es que, dado el supuesto de normalidad de los errores, será cierto que:

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

5. El Modelo en Desviaciones Respecto a la Media

El Modelo en Desviaciones Respecto a la Media

Definamos la siguiente matriz A :

$$A = I - i(i'i)^{-1}i' = I - i(n)^{-1}i' = I - \frac{1}{n}ii'$$

Donde i es un vector columna de unos. Esta matriz permite generar desviaciones respecto a la media.

Por ejemplo, si A premultiplica a un vector columna Y se obtiene:

$$AY = \left[I - \frac{1}{n}i'i \right] Y = Y - \frac{1}{n}i'iY = Y - i\bar{Y} = \begin{bmatrix} Y_1 - \bar{Y} \\ Y_2 - \bar{Y} \\ \vdots \\ Y_n - \bar{Y} \end{bmatrix}$$

El Modelo en Desviaciones Respecto a la Media

Para el modelo en desviaciones, se multiplica la matriz A al modelo estimado $Y = X\hat{\beta} + \hat{\varepsilon}$ y se obtiene:

$$AY = AX\hat{\beta} + A\hat{\varepsilon}$$

Particionamos a la matriz X en columna de unos, i , y el resto de variables explicativas en una matriz llamada X_2 , quedando:

$$AY = A[i \quad X_2] \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} + A\hat{\varepsilon}$$

En la última expresión $\hat{\beta}_2$ es un vector $(k - 1) \times 1$ de estimadores de las pendientes del modelo.

Dado que $A\hat{\varepsilon} = \hat{\varepsilon}$ (pues el promedio de $\hat{\varepsilon}$ es igual a cero) y como $Ai = 0$, resulta:

$$AY = AX_2\hat{\beta}_2 + \hat{\varepsilon}$$

Este es el modelo en desviaciones matricial, expresión similar al modelo en desviaciones bivariado, y donde no aparece el estimador del intercepto $\hat{\beta}_1$.

El Modelo en Desviaciones Respecto a la Media

De la expresión del modelo en desviaciones:

$$AY = AX_2\hat{\beta}_2 + \hat{\varepsilon}$$

Se puede deducir que el estimador MCO de las pendientes es:

$$\hat{\beta}_2 = [(AX_2)'AX]^{-1}(AX_2)'AY$$

$$\hat{\beta}_2 = [X_2'A'AX]^{-1}X_2'A'AY$$

$$\hat{\beta}_2 = (X_2'AX)^{-1}X_2'AY$$

6. Omisión de Variables Relevantes

Omisión de Variables Relevantes

Suponer que el modelo correctamente especificado es el siguiente:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

Entonces, si nosotros regresionamos Y sobre X_1 sin incluir X_2 , entonces el estimador es:

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y$$

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon)$$

$$\hat{\beta}_1 = \beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

Tomando esperanzas, vemos que si $X_1'X_2 = 0$ o $\beta_2 = 0$, $\hat{\beta}_1$ es insesgado. El resultado es:

$$E[\hat{\beta}_1|X] = \beta_1 + P_{12}\beta_2$$

Donde $P_{12} = (X_1'X_1)^{-1}X_1'X_2$. Existe sesgo por variable omitida.

Omisión de Variables Relevantes

Ejemplo

Tenemos el siguiente modelo con dos variables:

$$wage_i = \beta_1 + \beta_2 sch + \beta_3 abil + \varepsilon_i$$

Donde *abil* es la habilidad del individuo. Debido a que habilidad no es observada, se estima el siguiente modelo:

$$wage_i = \beta_1 + \beta_2 sch + u_i$$

Donde $u_i = \beta_3 abil + \varepsilon_i$. El estimador de β_2 de la regresión simple de *wage* sobre *sch* es etiquetada como $\tilde{\beta}_2$. La esperanza de $\tilde{\beta}_2$ es:

$$E[\tilde{\beta}_2 | sch_1, sch_2, \dots, sch_n] = \beta_2 + \beta_3 \frac{\sum_{i=1}^n (sch_i - \overline{sch}) abil_i}{\sum_{i=1}^n (sch_i - \overline{sch})^2}$$

Omisión de Variables Relevantes

El ratio multiplicando β_3 es el coeficiente pendiente de la regresión de *abil* y *sch*, que se puede escribir como:

$$abil_i = \tilde{\delta}_0 + \tilde{\delta}_1 sch_i$$

Asimismo, podemos escribir como:

$$E[\tilde{\beta}_2 | sch_1, sch_2, \dots, sch_n] = \beta_2 + \beta_3 \tilde{\delta}_1$$

Existen dos casos donde $\tilde{\beta}_2$ es insesgado: $\beta_3 = 0$ o $\tilde{\delta}_1 = 0$ incluso si $\beta_3 \neq 0$.

Cuando *sch* y *abil* están correlacionados, $\tilde{\delta}_1$ tiene el mismo signo como la correlación entre *sch* y *abil*

- $\tilde{\delta}_1 > 0$ si *sch* y *abil* están positivamente correlacionados.
- $\tilde{\delta}_1 < 0$ si *sch* y *abil* están negativamente correlacionados.

El signo del sesgo en $\tilde{\beta}_2$ depende de los signos de β_3 y $\tilde{\delta}_1$.

Omisión de Variables Relevantes

Para generalizar, asumimos el siguiente modelo:

$$Y_i = \beta_0 + \beta_1 X_1 + u_i$$

$$u_i = \beta_2 X_2 + \varepsilon_i$$

	$corr(X_1, X_2) > 0$	$corr(X_1, X_2) < 0$
$\beta_2 > 0$	Sesgo positivo	Sesgo negativo
$\beta_2 < 0$	Sesgo negativo	Sesgo positivo

7. Inclusión de Variables Irrelevantes

Inclusión de Variables Irrelevantes

Omitir variables relevantes puede generar sesgos en las estimaciones
¿qué ocurre si incluyen variables irrelevantes en el modelo?

El modelo correcto es $Y = X\beta + \varepsilon$, sin embargo el investigador propone el siguiente:

$$Y = X\beta + X_3\beta_3 + v$$

Esta ecuación se puede escribir como:

$$Y = X^*\beta^* + v$$

Donde $X^* = [X \quad X_3]$ y $\beta^{*'} = [\beta' \quad \beta_3']$. Entonces, las ecuaciones normales de la estimación MCO son:

$$(X^{*'}X^*)\tilde{\beta}^* = X^{*'}Y$$

Donde $\tilde{\beta}^{*'} = [\tilde{\beta}' \quad \tilde{\beta}_3']$ son los estimadores de los parámetros en el modelo incorrecto.

Inclusión de Variables Irrelevantes

Esta última ecuación se puede escribir como:

$$\begin{pmatrix} [X'] \\ [X'_3] \end{pmatrix} \begin{bmatrix} X & X_3 \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{\beta}_3 \end{bmatrix} = \begin{bmatrix} X' \\ X'_3 \end{bmatrix} Y$$

$$\begin{bmatrix} X'X & X'X_3 \\ X'_3X & X'_3X_3 \end{bmatrix} \begin{bmatrix} \tilde{\beta} \\ \tilde{\beta}_3 \end{bmatrix} = \begin{bmatrix} X' \\ X'_3 \end{bmatrix} Y$$

Expresando como dos ecuaciones:

$$(X'X)\tilde{\beta} + (X'X_3)\tilde{\beta}_3 = X'Y \quad (i)$$

$$(X'_3X)\tilde{\beta} + (X'_3X_3)\tilde{\beta}_3 = X'_3Y \quad (ii)$$

Despejando de (ii):

$$\tilde{\beta}_3 = (X'_3X_3)^{-1}[X'_3Y - X'_3X\tilde{\beta}]$$

Inclusión de Variables Irrelevantes

Reemplazando en (i)

$$(X'X)\tilde{\beta} + (X'X_3)(X_3'X_3)^{-1}[X_3'Y - X_3'X\tilde{\beta}] = X'Y$$

$$(X'X)\tilde{\beta} - (X'X_3)(X_3'X_3)^{-1}X_3'X\tilde{\beta} = X'Y - (X'X_3)(X_3'X_3)^{-1}X_3'Y$$

$$[X' - (X'X_3)(X_3'X_3)^{-1}X_3']X\tilde{\beta} = [X' - (X'X_3)(X_3'X_3)^{-1}X_3']Y$$

$$X'[I - X_3(X_3'X_3)^{-1}X_3']X\tilde{\beta} = X'[I - X_3(X_3'X_3)^{-1}X_3']Y$$

$$X'M_3X\tilde{\beta} = X'M_3Y$$

$$\tilde{\beta} = [X'M_3X]^{-1}X'M_3Y \quad (iii)$$

Donde $M_3 = I - X_3(X_3'X_3)^{-1}X_3'$ es la matriz generadora de residuos.

De manera similar, $\tilde{\beta}_3 = [X'MX]^{-1}X'MY$ donde $M = I - X(X'X)^{-1}X'$

Inclusión de Variables Irrelevantes

Evaluando insesgadez de β con la inclusión de las variables irrelevantes. Reemplazando el modelo correcto en (iii):

$$\tilde{\beta} = (X' M_3 X)^{-1} X' M_3 (X\beta + \varepsilon)$$

$$\tilde{\beta} = \beta + (X' M_3 X)^{-1} X' M_3 \varepsilon$$

Tomando el valor esperado, se comprueba que $\tilde{\beta}$ es insesgado dados los supuestos del modelo:

$$E[\tilde{\beta}] = \beta + (X' M_3 X)^{-1} X' M_3 E[\varepsilon]$$

$$E[\tilde{\beta}] = \beta$$

Cuando se agregan variables innecesarias al modelo, esto no produce ningún sesgo en las estimaciones.

Inclusión de Variables Irrelevantes

Evaluando la varianza de los estimadores. Definiendo la matriz de varianzas y covarianzas de $\tilde{\beta}$:

$$\begin{aligned} Var(\tilde{\beta}) &= E[(\tilde{\beta} - E[\tilde{\beta}])(\tilde{\beta} - E[\tilde{\beta}])'] \\ &= E[(X'M_3X)^{-1}X'M_3\varepsilon\varepsilon'M_3'X(X'M_3X)^{-1}] \\ &= (X'M_3X)^{-1}X'M_3E[\varepsilon\varepsilon']M_3'X(X'M_3X)^{-1} \end{aligned}$$

Como $E[\varepsilon\varepsilon'] = \sigma^2 I$ y M_3 es simétrica e idempotente, la matriz de varianzas y covarianzas resulta en:

$$Var(\tilde{\beta}) = \sigma^2$$

Comparando estos resultados con el estimador MCO del modelo correcto $\hat{\beta} = (X'X)^{-1}X'Y$, donde su varianza es $Var(\hat{\beta}) = \sigma^2(X'X)^{-1}$. Entonces:

$$Var(\hat{\beta}) < Var(\tilde{\beta})$$

En conclusión, incluir variables irrelevantes no sesga la estimación de β , pero sí incrementa la varianza de las estimaciones.

8. El Teorema de Frisch-Waugh

Teorema de Frisch-Waugh

Dado un modelo de regresión en matrices:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

Se puede obtener los estimadores de MCO de la pendiente de las variables X_1 , es decir $\hat{\beta}_1$, libre del efecto de las variables en X_2 en tres pasos como sigue:

1. Estimar $Y = X_2\beta_2 + v$ por MCO y calcular los residuos $\hat{\varepsilon}_Y = M_2Y$ donde $M_2 = I - X_2(X_2'X_2)^{-1}X_2'$
2. Estimar $X_1 = X_2\beta_2 + w$ por MCO y calcular los residuos $\hat{\varepsilon}_{X_1} = M_2X_1$
3. Estimar $\hat{\varepsilon}_Y = \hat{\varepsilon}_{X_1}\beta_1 + z$ por MCO. Este estimador es $\hat{\beta}_1 = (\hat{\varepsilon}_{X_1}'\hat{\varepsilon}_{X_1})^{-1}\hat{\varepsilon}_{X_1}'\hat{\varepsilon}_Y$

Reemplazando las expresiones de los residuos en $\hat{\beta}_1$ se obtiene:

$$\hat{\beta}_1 = (\hat{\varepsilon}_{X_1}'\hat{\varepsilon}_{X_1})^{-1}\hat{\varepsilon}_{X_1}'\hat{\varepsilon}_Y = (X_1'M_2M_2X_1)^{-1}X_1'M_2M_2Y$$

$$\hat{\beta}_1 = (X_1'M_2X_1)^{-1}X_1'M_2Y$$

Teorema de Frisch-Waugh

Reemplazando las expresiones de los residuos en $\hat{\beta}_1$ se obtiene:

$$\hat{\beta}_1 = (\hat{\varepsilon}'_{X_1} \hat{\varepsilon}_{X_1})^{-1} \hat{\varepsilon}'_{X_1} \hat{\varepsilon}_Y$$

$$\hat{\beta}_1 = (X'_1 M'_2 M_2 X_1)^{-1} X'_1 M'_2 M_2 Y$$

$$\hat{\beta}_1 = (X'_1 M_2 X_1)^{-1} X'_1 M_2 Y$$

De forma similar, el vector de parámetros $\hat{\beta}_2$ puede obtenerse mediante regresiones de residuos, siendo

$$\hat{\beta}_2 = (X'_2 M_1 X_2)^{-1} X'_2 M_1 Y$$

Donde $M_1 = I - X_1(X'_1 X_1)^{-1} X'_1$

La idea principal de este teorema es que, al regresionarse con los residuos, se han eliminado los efectos del resto de variables.

9. Descomposición de la Suma de Cuadrados

Descomposición de la Suma de Cuadrados

Calcularemos la suma de cuadrados totales $SCT = \sum(Y_i - \bar{Y})^2$. En términos matriciales, para obtener la sumatoria de cuadrados totales, tenemos que realizar el producto interno del vector AY .

De $AY = AX_2\hat{\beta}_2 + \hat{\varepsilon}$, luego:

$$Y'A'AY = (AX_2\hat{\beta}_2 + \hat{\varepsilon})'(AX_2\hat{\beta}_2 + \hat{\varepsilon})$$

$$Y'A'AY = \hat{\beta}_2'X_2'A'AX_2\hat{\beta}_2 + \hat{\beta}_2'X_2'A'\hat{\varepsilon} + \hat{\varepsilon}'AX_2\hat{\beta}_2 + \hat{\varepsilon}'\hat{\varepsilon}$$

Sabiendo que $A'\hat{\varepsilon} = \hat{\varepsilon}$ y que $X_2'\hat{\varepsilon} = 0$, esta ecuación se reduce a:

$$\underbrace{Y'A'AY}_{\text{Suma de cuadrados totales (SCT)}} = \underbrace{\hat{\beta}_2'X_2'A'AX_2\hat{\beta}_2}_{\text{Suma de cuadrados explicada (SCE)}} + \underbrace{\hat{\varepsilon}'\hat{\varepsilon}}_{\text{Suma de cuadrados de los residuos (SCR)}}$$

Descomposición de la Suma de Cuadrados

Estas expresiones tienen versiones equivalentes:

$$\text{Suma de cuadrados explicada (SCE)} \quad \Rightarrow \quad \hat{\beta}'_2 X'_2 A X_2 \hat{\beta}_2 \equiv \hat{\beta}'_2 X'_2 A Y = \hat{\beta}' X' X \hat{\beta} - n \bar{Y}^2$$

$$\text{Suma de cuadrados totales (SCT)} \quad \Rightarrow \quad Y' A Y \equiv Y' Y - n \bar{Y}^2$$

Demostración

Partiendo de $Y = X\hat{\beta} + \hat{\varepsilon}$, el producto interno es:

$$\begin{aligned} Y'Y &= (\hat{\varepsilon} + \hat{\beta}'X')(X\hat{\beta} + \hat{\varepsilon}) \\ Y'Y &= \hat{\beta}'X'X\hat{\beta} + \underbrace{\hat{\beta}'X'\hat{\varepsilon}}_{=0} + \underbrace{\hat{\varepsilon}X\hat{\beta}}_{X'\hat{\varepsilon}=0} + \hat{\varepsilon}'\hat{\varepsilon} \\ Y'Y &= \hat{\beta}'X'X\hat{\beta} + \hat{\varepsilon}'\hat{\varepsilon} \end{aligned}$$

Restamos $n\bar{Y}^2$ a ambos lados y obtenemos la SCT:

$$Y'Y - n\bar{Y}^2 = \hat{\beta}'X'X\hat{\beta} - n\bar{Y}^2 + \hat{\varepsilon}'\hat{\varepsilon}$$

Descomposición de la Suma de Cuadrados

Construyendo el **coeficiente de determinación** R^2 como:

$$R^2 = \frac{SCE}{SCT} = \frac{\hat{\beta}_2' X_2' AY}{Y' AY} = \frac{\hat{\beta}_2' X_2' A X_2 \hat{\beta}_2}{Y' AY} = 1 - \frac{\hat{\varepsilon}' \hat{\varepsilon}}{Y' AY}$$

Presenta el problema que siempre aumenta cuando se incorporan nuevas variables explicativas en la regresión.

El R^2 **ajustado** se define como:

$$R^2 \text{ ajustado} = 1 - \frac{\hat{\sigma}^2}{\hat{\sigma}_Y^2} = 1 - \frac{\frac{SCR}{n-K}}{\frac{SCT}{n-1}} = 1 - \left(\frac{n-1}{n-K} \right) \frac{SCR}{SCT}$$

Esta fórmula castiga la inclusión de muchas variables, en el sentido que si K aumenta, la SCR disminuye y paralelamente $\frac{n-1}{n-K}$ aumenta. Para que R^2 *ajustado* aumente, $SCR > \frac{n-1}{n-K}$, entonces la variable incluida si es relevante.

- **Obligatorio:**

Capítulo 1.1-1.3: Hayashi, Fumio (2000). Econometrics. *Princeton University Press*.

Capítulo 13: Gujarati, D. & Porter, D. (2010). Econometría. Quinta Edición. McGrawHill.

- **Opcional:**

Capítulo 4.1-4.3: Greene, Willian (2018). Econometrics Analysis (Eighth Edition). *Pearson: New York University*.

Capítulo 5.1-5.2: Hill, R. C., Griffiths, W. E., & Lim, G. C. (2018). *Principles of econometrics*. John Wiley & Sons.



PUCP