



Econometría I

El Modelo Lineal General (II): Contraste de hipótesis y previsión

Miguel Jerez y Sonia Sotoca

Universidad Complutense de Madrid

Septiembre 2009

Índice

- **Contraste de hipótesis**
- Estrategias de contraste
- Intervalos y regiones de confianza
- Previsión
- Estimación restringida

Contraste de hipótesis (I): Hipótesis generales-1

Se trata de contrastar el cumplimiento de m hipótesis lineales de la forma:

$$H_0 : \mathbf{A}\beta = \mathbf{c} \quad (1)$$

en donde \mathbf{A} es una matriz ($m \times k$) de números reales, de rango m , β es un vector ($k \times 1$) de valores de parámetros del MLG y \mathbf{c} es un vector ($m \times 1$) de números reales.

Para evaluar en qué medida los datos dan soporte a la hipótesis buscaremos una medida de distancia escalar, computable y que tenga una distribución conocida bajo H_0 .

Sea $\delta = \mathbf{A}\hat{\beta} - \mathbf{c}$, la desviación a que $\hat{\beta}$ da lugar con respecto al cumplimiento estricto de (1). Suponiendo que se cumplen las hipótesis del MLG, la distribución de δ bajo H_0 es:

$$\delta | \mathbf{X} \underset{H_0}{\sim} N[\mathbf{0}, \sigma_\varepsilon^2 \mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T]$$

este estadístico no es adecuado para contrastar (1) ya que: a) es un vector, lo que no permite contrastar conjuntamente todas las hipótesis, y b) su distribución depende de parámetros desconocidos. A partir de δ puede formarse un nuevo estadístico cuya distribución condicional es:

$$\bar{F} = \delta^T [\text{cov}(\delta)]^{-1} \delta = \frac{(\mathbf{A}\hat{\beta} - \mathbf{c})^T [\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A}\hat{\beta} - \mathbf{c})}{\sigma_\varepsilon^2} \underset{H_0}{\sim} \chi_m^2 \quad (2)$$

que se llama *distancia de Mahalanobis*. Este estadístico es una medida escalar de distancia al cumplimiento de la hipótesis nula, pero depende de un parámetro desconocido, σ_ε^2

Contraste de hipótesis (II): Hipótesis generales-1

Bajo las hipótesis del MLG puede demostrarse que:

$$(n-k) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2} \sim \chi_{n-k}^2 \quad (3)$$

y que esta variable aleatoria es independiente de (2). Por tanto dividiendo (2) y (3), y suponiendo que H_0 es cierta, se obtiene:

$$F = \frac{\frac{1}{m} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})^T [\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})}{\frac{1}{n-k} (n-k) \frac{\hat{\sigma}_\varepsilon^2}{\sigma_\varepsilon^2}} \sim F_{m,n-k}$$

o, equivalentemente:

$$F = \frac{(\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})^T [\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T]^{-1} (\mathbf{A}\hat{\boldsymbol{\beta}} - \mathbf{c})}{m \hat{\sigma}_\varepsilon^2} \underset{H_0}{\sim} F_{m,n-k} \quad (4)$$

El estadístico F es computable, ya que no depende de parámetros desconocidos, y escalar, por lo que mide la distancia total al cumplimiento conjunto de todas las hipótesis.

Contraste de hipótesis (III): Casos particulares-1

Si se desea contrastar una sola hipótesis, la expresión (1) puede escribirse como:

$$H_0 : \mathbf{a}^T \boldsymbol{\beta} = c$$

en donde \mathbf{a} es un vector $k \times 1$ de números reales y c es un escalar. En este caso, la expresión (4) se simplifica a:

$$F = \frac{(\mathbf{a}^T \hat{\boldsymbol{\beta}} - c)^2}{\hat{\sigma}_\varepsilon^2 \mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}} \underset{H_0}{\sim} F_{1, n-k}$$

y la raíz cuadrada de este estadístico se distribuye como una t de Student:

$$t = \frac{\mathbf{a}^T \hat{\boldsymbol{\beta}} - c}{\hat{\sigma}_\varepsilon \sqrt{\mathbf{a}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{a}}} \underset{H_0}{\sim} t_{n-k} \quad (5)$$

Un contraste muy importante de este tipo es el de significación individual de un parámetro, en donde la hipótesis nula es $H_0 : \beta_i = 0$ y, por tanto, el estadístico de contraste queda:

$$t = \frac{\hat{\beta}_i}{\text{s.d.}(\hat{\beta}_i)} \underset{H_0}{\sim} t_{n-k} \quad (6)$$

Contraste de hipótesis (IV): Casos particulares-2

A veces resulta interesante contrastar conjuntamente la hipótesis de que un grupo de parámetros sea estadísticamente distinto de cero. Por ejemplo, si se desea contrastar la significación de las h últimas componentes de β , la hipótesis nula se formularía como:

$$H_0 : \begin{bmatrix} \mathbf{0} & \mathbf{I}_h \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \mathbf{0}$$

donde β_1, β_2 son subvectores de β de dimensiones $(k-h)$ y h , respectivamente, e \mathbf{I}_h es la matriz identidad de dimensión $(h \times h)$. Particularizando (4) para este caso, se obtiene:

$$F = \frac{\hat{\beta}_2^T (\mathbf{V}_{22})^{-1} \hat{\beta}_2}{h \hat{\sigma}_\varepsilon^2} \sim F_{h, n-k}$$

en donde \mathbf{V}_{22} es la submatriz $(h \times h)$ rectangular inferior de $(\mathbf{X}^T \mathbf{X})^{-1}$ tal que $\hat{\sigma}_\varepsilon^2 \mathbf{V}_{22} = \text{côv}(\hat{\beta}_2)$

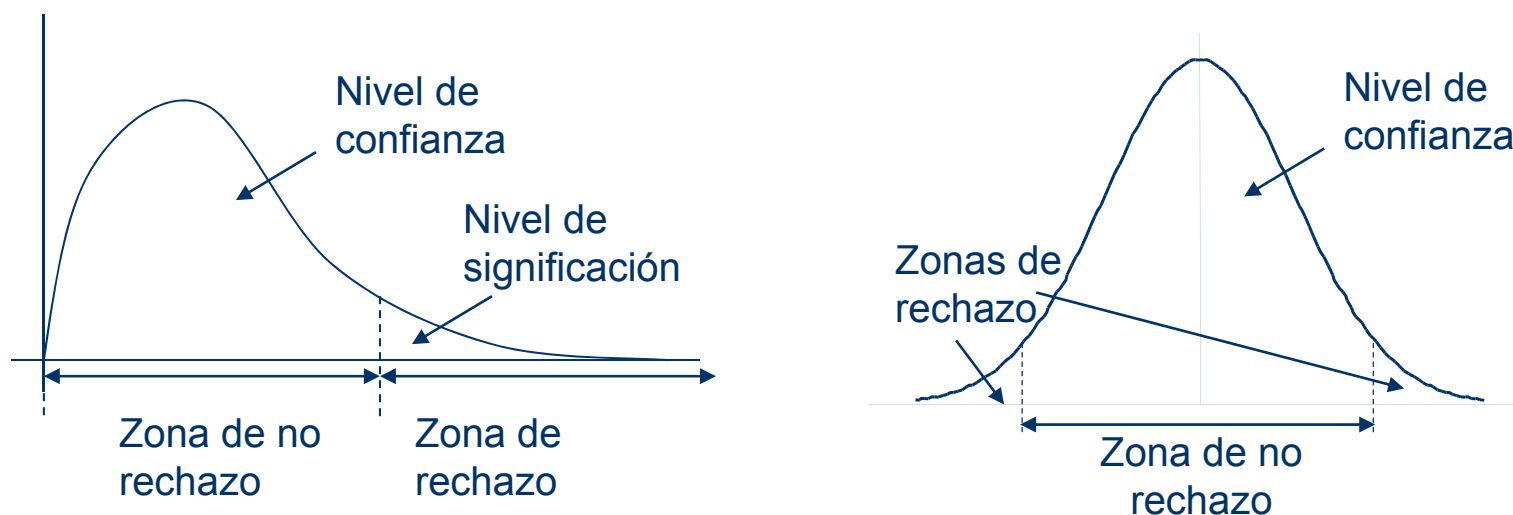
Índice

- Contraste de hipótesis
- **Estrategias de contraste**
- Intervalos y regiones de confianza
- Previsión
- Estimación restringida

Estrategias de contraste (I): nivel de significación

A partir de una medida de distancia escalar, computable y que tiene una distribución conocida bajo H_0 , pueden adoptarse dos estrategias básicas de contraste.

La primera consiste en **escoger a priori un nivel de significación para el contraste**, que mide el máximo riesgo que se está dispuesto a asumir de rechazar erróneamente H_0 .



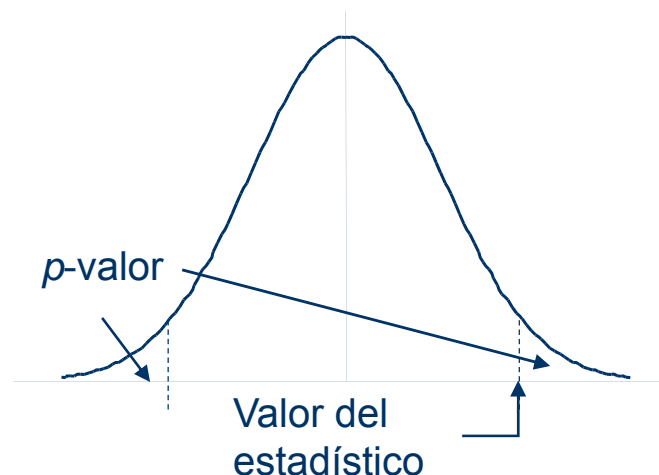
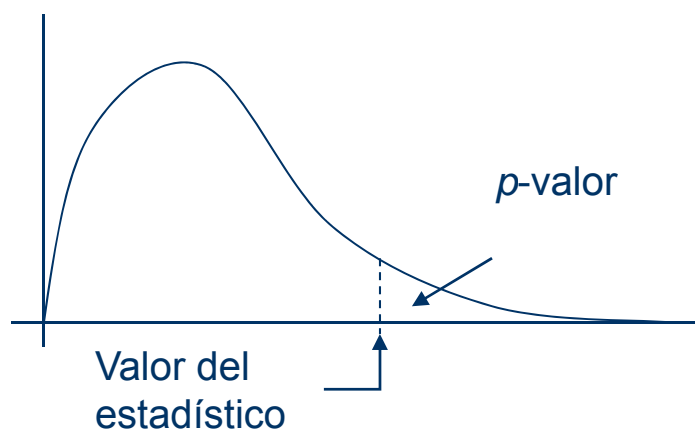
Elegir un nivel de significación concreto permite dividir el conjunto de valores posibles del estadístico en dos zonas: la zona de no rechazo y la zona de rechazo. Si el valor muestral del estadístico cae en la zona de rechazo, se decide que la hipótesis nula es falsa. En caso contrario, la muestra no permite rechazar la hipótesis. Debe tenerse en cuenta que **no rechazar H_0 no es lo mismo que aceptar H_0** .

La configuración de las zonas de rechazo y no rechazo depende de la distribución del estadístico bajo H_0 y del tipo de hipótesis (de una cola, de dos colas, ...).

Estrategias de contraste (II): p -valor

La segunda estrategia consiste en **medir el nivel de significación marginal (o p -valor) del estadístico**.

El p -valor mide el riesgo de rechazar erróneamente H_0 , dado el valor que se ha obtenido del estadístico. Por tanto, cuanto menor sea el p -valor, más evidencia tenemos en contra de la hipótesis nula y más tenderemos a rechazarla.



Ambas estrategias (escoger *a priori* un nivel de significación para el contraste y decidir a partir del p -valor) son válidas. Suele usarse más el p -valor, porque es menos arbitrario y, además, los programas econométricos modernos suelen calcularlo de forma automática.

Índice

- Contraste de hipótesis generales
- Estrategias de contraste
- **Intervalos y regiones de confianza**
- Previsión
- Estimación restringida

Intervalos y regiones de confianza (I)

A veces interesa calcular entre qué valores pueden oscilar los parámetros para un nivel de confianza dado. Además de su propio interés, los *intervalos de confianza* o *regiones de confianza* resultantes pueden usarse como una estrategia alternativa de contraste de hipótesis.

Para calcular intervalos de confianza necesitamos relacionar los parámetros desconocidos con valores computables, mediante una expresión con una distribución conocida bajo hipótesis estándar. Bajo las hipótesis del MLG, se cumple que: $\hat{\beta} \sim N\left[\beta, \sigma_\varepsilon^2 (\mathbf{X}^T \mathbf{X})^{-1}\right]$ y, por tanto:

$$\frac{(\hat{\beta} - \beta)^T \mathbf{X}^T \mathbf{X} (\hat{\beta} - \beta)}{\sigma_\varepsilon^2} \sim \chi_k^2$$

el denominador de este estadístico no puede evaluarse, ya que la varianza del error es habitualmente desconocida. Utilizando de nuevo (3) y teniendo en cuenta que ambas variables son independientes, llegamos al siguiente resultado:

$$\frac{(\hat{\beta} - \beta)^T \mathbf{X}^T \mathbf{X} (\hat{\beta} - \beta)}{k \hat{\sigma}_\varepsilon^2} \sim F_{k, n-k} \quad (7)$$

consecuentemente, si se fija un nivel de significación arbitrario α , y teniendo en cuenta que la distribución F de Snedecor es siempre positiva, el siguiente resultado es cierto:

$$P\left[(\hat{\beta} - \beta)^T \mathbf{X}^T \mathbf{X} (\hat{\beta} - \beta) \leq k \hat{\sigma}_\varepsilon^2 F_{k, n-k}(\alpha)\right] = 1 - \alpha$$

siendo $F_{k, n-k}(\alpha)$ el percentil $1-\alpha$ de una variable F de Snedecor de k y $n-k$ grados de libertad.

Intervalos y regiones de confianza (II)

El argumento de $P[\]$ en la expresión anterior delimita un subconjunto de \mathbb{R}^k , que es la región de confianza que buscamos. La derivación de un resultado análogo para un subconjunto de los parámetros del modelo resulta inmediata.

Asimismo, mediante un desarrollo análogo al anterior se obtiene que:

$$\frac{\hat{\beta}_i - \beta_i}{\text{s.d.}(\hat{\beta}_i)} \sim t_{n-k} \quad (8)$$

por tanto, fijando un nivel de significación arbitrario α , se obtiene un intervalo de confianza para el verdadero valor de un sólo parámetro:

$$P\left[\hat{\beta}_i - t_{n-k}(\alpha/2) \text{s.d.}(\hat{\beta}_i) \leq \beta_i \leq \hat{\beta}_i + t_{n-k}(\alpha/2) \text{s.d.}(\hat{\beta}_i)\right] = 1 - \alpha$$

siendo $t_{n-k}(\alpha/2)$ el percentil $1 - \alpha/2$ de una variable t de Student de $n-k$ grados de libertad.

Índice

- Contraste de hipótesis generales
- Estrategias de contraste
- Intervalos y regiones de confianza
- **Previsión**
- Estimación restringida

Previsión (I): Previsión puntual

Sea un modelo de regresión lineal:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (9)$$

Una vez estimado el modelo para una muestra de tamaño n , nos planteamos el problema de estimar las m siguientes observaciones de la variable endógena, suponiendo que:

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \quad (10)$$

en donde:

\mathbf{y}^* : Vector ($m \times 1$) con m valores desconocidos de la variable endógena.

\mathbf{X}^* : Matriz ($m \times k$) con m valores conocidos de las k variables exógenas del modelo.

$\boldsymbol{\varepsilon}^*$: Vector ($m \times 1$) de perturbaciones aleatorias, cuyas propiedades estocásticas son idénticas a las del vector $\boldsymbol{\varepsilon}$ en (9)

Utilizaremos como predictor la esperanza de \mathbf{y}^* condicionada a \mathbf{X}^* y a la estimación de $\boldsymbol{\beta}$ obtenida por MCO a partir de la muestra (n observaciones). Formalmente:

$$E(\mathbf{y}^* | \mathbf{X}^*, \hat{\boldsymbol{\beta}}) = \hat{\mathbf{y}}^* = \mathbf{X}^* \hat{\boldsymbol{\beta}} \quad (11)$$

Bajo los supuestos habituales (excepto normalidad) el predictor (11) tiene dos propiedades fundamentales: 1) la esperanza del error de previsión es nula y 2) su varianza es mínima dentro de la familia de predictores lineales con esperanza nula del error.

Previsión (II): Propiedades de la previsión puntual

1) **Esperanza del error de previsión.** Si definimos el error de previsión como:

$$\tilde{\mathbf{y}}^* = \hat{\mathbf{y}}^* - \mathbf{y}^* \quad (12)$$

restando (11) y (10), se obtiene:

$$\hat{\mathbf{y}}^* - \mathbf{y}^* = \mathbf{X}^* (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \boldsymbol{\varepsilon}^*$$

por tanto, si $\hat{\boldsymbol{\beta}}$ es un estimador insesgado, al tomar esperanzas a ambos lados de (12) resulta:

$$E(\tilde{\mathbf{y}}^*) = \mathbf{0}$$

2) **Eficiencia.** En primer lugar obtendremos la expresión de la varianza del error de previsión. Por definición:

$$\begin{aligned} \text{cov}(\tilde{\mathbf{y}}^*) &= E(\tilde{\mathbf{y}}^* \tilde{\mathbf{y}}^{*T}) = E\left\{ \left[\mathbf{X}^* (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \boldsymbol{\varepsilon}^* \right] \left[\mathbf{X}^* (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) - \boldsymbol{\varepsilon}^* \right]^T \right\} \\ &= E\left[\mathbf{X}^* (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^{*T} + \boldsymbol{\varepsilon}^* \boldsymbol{\varepsilon}^{*T} - \mathbf{X}^* (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \boldsymbol{\varepsilon}^{*T} - \boldsymbol{\varepsilon}^* (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})^T \mathbf{X}^{*T} \right] \end{aligned}$$

Por la hipótesis de ausencia de autocorrelación, los errores en (9) y (10) son independientes y las esperanzas de los productos cruzados en la expresión anterior son nulas, de donde resulta:

$$\text{cov}(\tilde{\mathbf{y}}^*) = \sigma^2 \left[\mathbf{I} + \mathbf{X}^* (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^{*T} \right] \quad (13)$$

Si en vez de (11) se utilizara un predictor lineal alternativo, $\hat{\mathbf{y}}^* = \mathbf{X}^* \hat{\boldsymbol{\beta}}$ la matriz de covarianzas de la predicción sería: $\text{cov}(\hat{\mathbf{y}}^* - \mathbf{y}^*) = \sigma^2 \mathbf{I} + \mathbf{X}^* \text{cov}(\hat{\boldsymbol{\beta}}) \mathbf{X}^{*T}$ y, por el teorema de Gauss-Markov $\text{cov}(\hat{\mathbf{y}}^* - \mathbf{y}^*) - \text{cov}(\tilde{\mathbf{y}}^*)$ es una matriz semidefinida positiva.

Previsión (III): Previsión por intervalo

Bajo las hipótesis habituales, $\tilde{\mathbf{y}}^*$ es una combinación lineal de variables aleatorias normales, por lo que:

$$\tilde{\mathbf{y}}^* \sim N\left\{ \mathbf{0}, \sigma^2 \left[\mathbf{I} + \mathbf{X}^* (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^{*T} \right] \right\} \quad \frac{\tilde{\mathbf{y}}^{*T} \left[\mathbf{I} + \mathbf{X}^* (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^{*T} \right]^{-1} \tilde{\mathbf{y}}^*}{\sigma^2} \sim \chi_m^2$$

en donde el denominador de la última expresión es desconocido. Teniendo en cuenta nuevamente el resultado (3) y la independencia de ambas distribuciones, se cumplirá que:

$$\frac{\tilde{\mathbf{y}}^{*T} \left[\mathbf{I} + \mathbf{X}^* (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^{*T} \right]^{-1} \tilde{\mathbf{y}}^*}{m \hat{\sigma}^2} \sim F_{m, n-k}$$

Por tanto, si se fija un nivel de significación arbitrario α , se tiene que:

$$P \left\{ \tilde{\mathbf{y}}^{*T} \left[\mathbf{I} + \mathbf{X}^* (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^{*T} \right]^{-1} \tilde{\mathbf{y}}^* \leq m \hat{\sigma}^2 F_{m, n-k}(\alpha) \right\} = 1 - \alpha$$

siendo $F_{m, n-k}(\alpha)$ el percentil $1 - \alpha$ de una variable F de Snedecor de m y $n-k$ grados de libertad. En el caso particular de la previsión un período hacia adelante, tendríamos que:

$$\frac{\hat{y}_{n+1} - y_{n+1}}{\hat{\sigma} \sqrt{1 + \mathbf{x}_{n+1}^{*T} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_{n+1}^*}} = \frac{\hat{y}_{n+1} - y_{n+1}}{\text{s.d.}(\hat{y}_{n+1})} \sim t_{n-k}$$

$$\Rightarrow P \left[\hat{y}_{n+1} - t_{n-k}(\alpha/2) \text{s.d.}(\hat{y}_{n+1}) \leq y_{n+1} \leq \hat{y}_{n+1} + t_{n-k}(\alpha/2) \text{s.d.}(\hat{y}_{n+1}) \right] = 1 - \alpha$$

siendo $t_{n-k}(\alpha/2)$ el percentil $1 - \alpha/2$ de una variable t de Student de $n-k$ grados de libertad.

Índice

- Contraste de hipótesis generales
- Estrategias de contraste
- Intervalos y regiones de confianza
- Previsión
- **Estimación restringida**

Estimación restringida (I)

En ocasiones, puede resultar interesante estimar un modelo de regresión sujeto a restricciones lineales del tipo (1), porque:

- un análisis previo indica que estas restricciones son plausibles e imponerlas permite obtener estimaciones más precisas, o bien,
- se desea contrastar si la muestra rechaza estas restricciones, comparando los resultados que se obtienen estimando libremente y con restricciones

Las restricciones lineales pueden imponerse:

- por sustitución, o
- aplicando el método de mínimos cuadrados restringidos linealmente (MCRL).

Ejemplo (estimación restringida mediante sustitución): Sea la siguiente función Cobb-Douglas transformada logarítmicamente:

$$\ln y_t = \beta_0 + \beta_1 \ln k_t + \beta_2 \ln l_t + \varepsilon_t \quad (t = 1, 2, \dots, n)$$

... y supongamos que se quiere imponer la restricción $\beta_1 + \beta_2 = 1$ (rendimientos constantes de escala). Esto puede conseguirse reescribiendo el modelo como:

$$\ln y_t = \beta_0 + \beta_1 \ln k_t + (1 - \beta_1) \ln l_t + \varepsilon_t ; \quad \ln y_t - \ln l_t = \beta_0 + \beta_1 (\ln k_t - \ln l_t) + \varepsilon_t ;$$

$$\ln \frac{y_t}{l_t} = \beta_0 + \beta_1 \ln \frac{k_t}{l_t} + \varepsilon_t$$

Por tanto, la restricción puede imponerse eliminando un parámetro y transformando las variables, lo que da lugar a un nuevo modelo que puede estimarse libremente por MCO.

Estimación restringida (II): Estimador MCRL

Asimismo, se pueden imponer restricciones generales sobre las estimaciones resolviendo el problema:

$$\text{Min } S(\hat{\beta}) = (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta})$$

$$\text{sujeto a: } \mathbf{A}\hat{\beta} = \mathbf{c}$$

... cuyo Lagrangiano es: $\mathcal{L}(\hat{\beta}, \mu) = (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta}) + \mu^T (\mathbf{A}\hat{\beta} - \mathbf{c})$ o, equivalentemente:

$$\begin{aligned} \mathcal{L}(\hat{\beta}, \lambda) &= (\mathbf{y} - \mathbf{X}\hat{\beta})^T (\mathbf{y} - \mathbf{X}\hat{\beta}) + 2\lambda^T (\mathbf{A}\hat{\beta} - \mathbf{c}) \\ &= \mathbf{y}^T \mathbf{y} - 2\hat{\beta}^T \mathbf{X}^T \mathbf{y} + \hat{\beta}^T \mathbf{X}^T \mathbf{X} \hat{\beta} + 2\lambda^T \mathbf{A}\hat{\beta} - 2\lambda^T \mathbf{c} \\ &= \mathbf{y}^T \mathbf{y} - 2\hat{\beta}^T \mathbf{X}^T \mathbf{y} + \hat{\beta}^T \mathbf{X}^T \mathbf{X} \hat{\beta} + 2\hat{\beta}^T \mathbf{A}^T \lambda - 2\mathbf{c}^T \lambda \end{aligned}$$

Las condiciones de primer orden para la resolución del problema son:

$$\frac{\partial \mathcal{L}(\hat{\beta})}{\partial \hat{\beta}^T} = \mathbf{0}; -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X} \hat{\beta}_{MCRL} + 2\mathbf{A}^T \lambda_{MCRL} = \mathbf{0}; \hat{\beta}_{MCRL} = \hat{\beta}_{MCO} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \lambda_{MCRL}$$

$$\frac{\partial \mathcal{L}(\lambda)}{\partial \lambda} = \mathbf{0}; \mathbf{A}\hat{\beta}_{MCRL} = \mathbf{c}$$

sustituyendo la primera condición en la segunda se obtiene:

$$\mathbf{A}\hat{\beta}_{MCO} - \mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \lambda_{MCRL} = \mathbf{c}; \lambda_{MCRL} = \left[\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \right]^{-1} (\mathbf{A}\hat{\beta}_{MCO} - \mathbf{c})$$

y sustituyendo de nuevo este resultado en la primera condición resulta:

$$\hat{\beta}_{MCRL} = \hat{\beta}_{MCO} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \left[\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \right]^{-1} (\mathbf{A}\hat{\beta}_{MCO} - \mathbf{c}) \quad (14)$$

Estimación restringida (III): Propiedades del estimador MCRL

Las propiedades del estimador restringido, en comparación con las del MCO, son:

- Si las restricciones son ciertas y el estimador MCO es insesgado, el estimador MCRL también es insesgado. Análogamente, si las restricciones son falsas, el estimador MCRL es sesgado.
- La matriz $\text{COV}(\hat{\beta}_{MCO}) - \text{COV}(\hat{\beta}_{MCRL})$ es semidefinida positiva, por lo que el estimador MCRL tiene menos incertidumbre que el MCO y, consecuentemente,
- si las restricciones son ciertas, el estimador MCRL es más eficiente que el MCO†.
- La suma de cuadrados de los residuos del estimador MCRL es mayor o igual que la del estimador MCO: $S(\hat{\beta}_{MCRL}) \geq S(\hat{\beta}_{MCO})$

Esta última propiedad sugiere que la contrastación de hipótesis podría reducirse a estudiar si imponer la hipótesis como restricción cambia significativamente el ajuste del modelo.

Concretamente, si $H_0 : \mathbf{A}\beta = \mathbf{c}$ es cierta, el contraste habitual de hipótesis puede realizarse a partir de los estadísticos:

$$F = \frac{n-k}{m} \frac{S(\hat{\beta}_{MCRL}) - S(\hat{\beta}_{MCO})}{S(\hat{\beta}_{MCO})} \underset{H_0}{\sim} F_{m, n-k} \quad (15)$$

$$F = \frac{n-k}{m} \frac{R_{MCO}^2 - R_{MCRL}^2}{1 - R_{MCO}^2} \underset{H_0}{\sim} F_{m, n-k} \quad (16)$$

† Esto no contradice el Teorema de Gauss-Markov, ya que MCRL utiliza más información (las restricciones) que MCO.

Estimación restringida (IV): Demostraciones

Para demostrar que (15) es equivalente a (4) partimos de:

$$\begin{aligned}\hat{\varepsilon}_{MCRL} &= \mathbf{y} - \mathbf{X}\hat{\beta}_{MCRL} = \mathbf{y} - \mathbf{X}\hat{\beta}_{MCRL} + \mathbf{X}\hat{\beta}_{MCO} - \mathbf{X}\hat{\beta}_{MCO} = \\ &= \mathbf{y} - \mathbf{X}\hat{\beta}_{MCO} - \mathbf{X}(\hat{\beta}_{MCRL} - \hat{\beta}_{MCO}) = \hat{\varepsilon}_{MCO} - \mathbf{X}(\hat{\beta}_{MCRL} - \hat{\beta}_{MCO})\end{aligned}$$

Por tanto, denotando $S(\hat{\beta}_{MCRL}) = \hat{\varepsilon}_{MCRL}^T \hat{\varepsilon}_{MCRL}$; $S(\hat{\beta}_{MCO}) = \hat{\varepsilon}_{MCO}^T \hat{\varepsilon}_{MCO}$ y teniendo en cuenta que, por las propiedades de MCO, las variables explicativas son ortogonales a los residuos:

$$S(\hat{\beta}_{MCRL}) = S(\hat{\beta}_{MCO}) + (\hat{\beta}_{MCRL} - \hat{\beta}_{MCO})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{MCRL} - \hat{\beta}_{MCO});$$

$$S(\hat{\beta}_{MCRL}) - S(\hat{\beta}_{MCO}) = (\hat{\beta}_{MCRL} - \hat{\beta}_{MCO})^T \mathbf{X}^T \mathbf{X} (\hat{\beta}_{MCRL} - \hat{\beta}_{MCO}); \text{ por (14)}$$

$$S(\hat{\beta}_{MCRL}) - S(\hat{\beta}_{MCO}) =$$

$$= (\mathbf{A}\hat{\beta}_{MCO} - \mathbf{c})^T \left[\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \right]^{-1} \mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \left[\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \right]^{-1} (\mathbf{A}\hat{\beta}_{MCO} - \mathbf{c})$$

$$= (\mathbf{A}\hat{\beta}_{MCO} - \mathbf{c})^T \left[\mathbf{A}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{A}^T \right]^{-1} (\mathbf{A}\hat{\beta}_{MCO} - \mathbf{c})$$

por tanto, la diferencia de sumas residuales coincide con el numerador del estadístico F habitual. En cuanto al denominador, es trivial que:

$$m\hat{\sigma}_\varepsilon^2 = \frac{m}{n-k} S(\hat{\beta}_{MCO})$$

La expresión (16) se obtiene inmediatamente teniendo en cuenta que, por definición:

$$R_{MCO}^2 = 1 - \frac{S(\hat{\beta}_{MCO})}{(\mathbf{y} - \hat{\mu}_y)^T (\mathbf{y} - \hat{\mu}_y)}; R_{MCRL}^2 = 1 - \frac{S(\hat{\beta}_{MCRL})}{(\mathbf{y} - \hat{\mu}_y)^T (\mathbf{y} - \hat{\mu}_y)}$$

Miguel Jerez (mjerez@ccee.ucm.es)

Sonia Sotoca (sotoca@ccee.ucm.es)

Departamento de Fundamentos del Análisis Económico II
(Economía Cuantitativa)

Facultad de Ciencias Económicas, UCM

Más materiales en:

<http://www.ucm.es/info/ecocuan/mjm/ctr1mj>

<http://econometriamj.blogspot.com/>