MAGCEA Econometria IN709 - Otoño 2008
Universidad de Chile
Departamento de Ingeniería Industrial
Profesor: Mattia Makovec (mmakovec@dii.uchile.cl)
Auxiliares: Felipe Avilés, Gonzalo Valdés

Clase 2

12/3/2008

(Versión preliminar, no difundir)

Estimador de mínimos cuadrados ordinarios (MCO) con regresores no estocásticos

Dado el modelo lineal general $y = X\beta + \varepsilon$ basado en una muestra de tamaño N, suponemos que se cumplan los supuestos basicos H1-H6 ilustrados en la clase anterior:

H1 Linealidad: la variable dependiente y es una función lineal del vector de parametros β; también se supone que los parametros $\beta_1, \beta_2....\beta_k$ sean constantes por todas la N observaciones en la muestra.

H2 Condición de identificación: X es una matriz de rango completo de columnas y $rango(X) = k$.

H3 $E(\varepsilon_i) = 0 \; \forall i = 1...N$, o en forma vectorial: $E(\varepsilon) = 0$.

H4 $Var(\varepsilon_i) = \sigma^2 \; \forall i = 1...N$, $Cov(\varepsilon_i \varepsilon_j) = 0 \; \forall i \neq j$, o en forma matricial: $E(\varepsilon\varepsilon') = \sigma^2 I_N$ (matriz de varianza y covarianza escalar).

H5 X es una matriz de regresores no estocásticos.

H6 Los errores siguen una distribución normal con media 0 y varianza constante σ^2: $\varepsilon_i \sim N(0, \sigma^2) \; \forall i = 1...N$; $\varepsilon \sim N(0, \sigma^2 I_N)$.

Se analiza como obtener el estimador de mínimos cuadrados ordinarios. El problema se traduce en minimizar la suma cuadratica residual:

$$Min \colon \varepsilon'\varepsilon$$

sustituyendo $\varepsilon = y - X\beta$ el problema se puede formular como una minimización de la siguiente función lagrangeana $L(\beta)$ respecto a β :

$$\begin{aligned}
\underset{\beta}{Min} : \; L(\beta) &= (y - X\beta)'(y - X\beta) \\
\underset{\beta}{Min} : \; L(\beta) &= y'y - y'X\beta - \beta'X'y + \beta'X'X\beta \\
\underset{\beta}{Min} : \; L(\beta) &= y'y - 2\beta'X'y + \beta'X'X\beta
\end{aligned}$$

Dado que los terminos $y'X\beta$ y $\beta'X'y$ son el mismo escalar, entonces $(-y'X\beta - \beta'X'y) = -2\beta'X'y$. La condición del primer orden es:

$$\left[\frac{\partial L(\beta)}{\partial \beta}\right]_{\beta = \hat{\beta}_{MCO}} = 0$$

y se obtiene:

$$\begin{aligned}
-2X'y + 2X'X\hat{\beta}_{MCO} &= 0 \\
(X'X)\hat{\beta}_{MCO} &= X'y \quad \text{(ecuaciones normales)} \\
\hat{\beta}_{MCO} &= (X'X)^{-1}X'y
\end{aligned}$$

La matriz $(k \times k)$ $X'X$ es invertible dado que por el supuesto H2, X es invertible, siendo una matriz de rango completo de columnas. Para averiguar que $\hat{\beta}_{MCO}$ sea efectivamente el mínimo de la función Lagrangeana $L(\beta)$, tenemos que comprobar que la derivada segunda $\frac{\partial^2 L(\beta)}{\partial\beta\partial\beta'}$ sea una matriz definida positiva.

$$\left[\frac{\partial^2 L(\beta)}{\partial\beta\partial\beta'}\right]_{\beta=\hat{\beta}_{MCO}} = \frac{\partial(-2X'y + 2X'X\hat{\beta}_{MCO})}{\partial\beta} = 2X'X$$

Consideramos una cualquiera transformación de la matriz $X'X$: $q = c'X'Xc$ donde c es un vector $(k \times 1)$ y $c \neq 0$. Si se define $v = Xc$, entonces $q = v'v = \sum_{i=1}^{N} v_i^2$. Se puede observar que, a menos que todos los elementos del vector v sean iguales a 0, q es siempre un escalar postivo, siendo una suma de terminos al cuadrado. Por otro lado, no puede ser que todos los elementos de v sean iguales a 0, dado que eso implicaria que algunas de las columnas de X fueran linealmente dependientes, siendo $c \neq 0$, violando el supuesto H2. Entonces, si $q = c'X'Xc$ es siempre positivo por cualquier $c \neq 0$, también $q = X'X$ con $c' = \underbrace{[1\ 1\ 1\ .....1]}_{(k\times 1)}$ será una matriz definida positiva.

Sustituyendo $y = X\beta + \varepsilon$ se obtiene una expresión alternativa de $\hat{\beta}_{MCO}$

$$\begin{aligned}
\hat{\beta}_{MCO} &= (X'X)^{-1}X'(X\beta + \varepsilon) \\
&= \underbrace{(X'X)^{-1}X'X}_{I}\beta + (X'X)^{-1}X'\varepsilon \\
&= \beta + (X'X)^{-1}X'\varepsilon
\end{aligned}$$

Utlizando la expresión $(\hat{\beta}_{MCO} - \beta) = (X'X)^{-1}X'\varepsilon$ se determina la **matriz de varianza y covarianza de** $\hat{\beta}_{MCO}$:

$$\begin{aligned}
Var(\hat{\beta}_{MCO}) &= E[(\hat{\beta}_{MCO} - E(\hat{\beta}_{MCO}))(\hat{\beta}_{MCO} - E(\hat{\beta}_{MCO}))'] \\
&= E[(\hat{\beta}_{MCO} - \beta)(\hat{\beta}_{MCO} - \beta)'] \quad \text{sustituyendo } \hat{\beta}_{MCO} - \beta = (X'X)^{-1}X'\varepsilon \\
&= E[((X'X)^{-1}X'\varepsilon)((X'X)^{-1}X'\varepsilon)'] \\
&= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'\underbrace{E(\varepsilon\varepsilon')}_{=\sigma^2 I}X(X'X)^{-1} \\
&= (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1} \\
&= \sigma^2(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{=I} \\
&= \sigma^2\underset{(k\times k)}{(X'X)^{-1}}
\end{aligned}$$

Propiededades del estimador de mínimos cuadrados ordinarios $\hat{\beta}_{MCO}$ en muestras finitas.

Se demuestra que bajo los supuestos H1-H5 el estimador de mínimos cuadrados ordinarios de β es lineal, insesgado y de varianza mínima entre los estimadores lineales y insesgados. Bajo el supuesto adicional H6 de normalidad de los errores, se demuestra también que $\hat{\beta}_{MCO}$ tiene una distribución normal multivariante: $\hat{\beta}_{MCO} \sim N(\beta, \sigma^2(X'X)^{-1})$.

- **Linealidad:** $\hat{\beta}_{MCO}$ es lineal en β entonces es lineal en y y en ε.

- **Insesgadez:** $E(\hat{\beta}_{MCO}) = \beta$

$$\begin{aligned} E(\hat{\beta}_{MCO}) &= E[\beta + (X'X)^{-1}X'\varepsilon] \\ &= \underbrace{E(\beta)}_{=\beta} + (X'X)^{-1}X'\underbrace{E(\varepsilon)}_{=0} \\ &= \beta \quad \hat{\beta}_{MCO} \text{ es estimador insesgado de } \beta. \end{aligned}$$

- **Eficiencia (Teorema de Gauss-Markov):**

Bajo los supuestos H1-H5, el estimador de mínimos cuadrados ordinarios $\hat{\beta}_{MCO}$ es el estimador más eficiente (o optimo), en el sentido de la varianza mínima, entre la clase de los estimadores lineales y insesgados de β.

Prueba:

Se considera un cualquier estimador lineal y insesgado de β, definido $\tilde{\beta} = Cy$, donde C es una matriz $(k \times N)$ que determina una cualquiera transformación lineal del vector y. El hecho que Cy sea estimador insesgado de β implica la siguiente propiedad de la matriz C:

$$\begin{aligned} E(\tilde{\beta}) &= E(Cy) = \beta \\ E(C(X\beta + \varepsilon)) &= \beta \\ E(CX\beta + C\varepsilon) &= \beta \\ CX\beta + C\underbrace{E(\varepsilon)}_{=0} &= \beta \\ CX\beta &= \beta \\ CX &= I \end{aligned}$$

Dado que $CX = I_k$, el estimador $\tilde{\beta} = Cy$ se puede expresar como: $\tilde{\beta} = C(X\beta + \varepsilon) = CX\beta + C\varepsilon = \beta + C\varepsilon$. Utilizando la propiedad $(\tilde{\beta} - \beta) = C\varepsilon$ se determina la varianza de $\tilde{\beta}$:

$$\begin{aligned} Var(\tilde{\beta}) &= E[(\tilde{\beta} - E(\tilde{\beta}))(\tilde{\beta} - E(\tilde{\beta}))'] \\ &= E[(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'] \\ &= E[(C\varepsilon)(C\varepsilon)'] \\ &= E[C\varepsilon\varepsilon'C'] \\ &= CE(\varepsilon\varepsilon')C' \\ &= C\sigma^2 IC' \\ &= \sigma^2 CC' \end{aligned}$$

Para calcular CC' se define la matriz $(k \times N)$ D tal que:

$$\begin{aligned} D &= C - (X'X)^{-1}X' \\ C &= D + (X'X)^{-1}X' \end{aligned}$$

Se observa también que $DX = 0 : DX = [C - (X'X)^{-1}X']X = CX - (X'X)^{-1}X'X = I_k - I_k = 0$. Sustituyendo la expresión de C en la de $Var(\tilde{\beta})$ se obtiene:

$$\begin{aligned} Var(\tilde{\beta}) &= \sigma^2 CC' = [D + (X'X)^{-1}X'][D + (X'X)^{-1}X']' \\ &= \sigma^2[D + (X'X)^{-1}X'][D' + X(X'X)^{-1}] \\ &= \sigma^2[DD' + (X'X)^{-1}X'D' + DX(X'X)^{-1} + (X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{=I_k}] \\ &= \sigma^2[DD' + (X'X)^{-1}(\underbrace{DX}_{=0})' + \underbrace{DX}_{=0}(X'X)^{-1} + (X'X)^{-1}] \\ &= \sigma^2[DD' + (X'X)^{-1}] \\ &= \sigma^2 DD' + \sigma^2(X'X)^{-1} \\ &= \sigma^2 DD' + var(\hat{\beta}_{MCO}) \end{aligned}$$

Entonces se obtiene que la diferencia entre la matriz de varianza y covarianza de $\tilde{\beta}$ y la de $\hat{\beta}_{MCO}$ es la matriz $(k \times k)$ $\sigma^2 DD'$ definida positiva, y se prueba que la varianza de $\hat{\beta}_{MCO}$ es la mínima entre los estimadores lineales y insesgados de β :

$$Var(\tilde{\beta}) - var(\hat{\beta}_{MCO}) = \sigma^2 DD' \qquad \text{definida positiva}$$

El estimador de mínimos cuadrados ordinarios $\hat{\beta}_{MCO}$ se define entonces el estimador linear insesgado optimo (abreviado en español con el acronimo E.L.I.O.) o el mejor estimador lineal insesgado (M.E.L.I.) en el sentido de la varianza mínima (en inglés el acronimo utilizado es B.L.U.E.: best linear unbiased estimator).

- **Normalidad de $\hat{\beta}_{MCO}$:** la linealidad, la insesgadez y la eficencia del estimador $\hat{\beta}_{MCO}$ no depienden sobre el supuesto H6 de normalidad de los errores. Bajo el supuesto adicional de errores normales, se establece que el estimador $\hat{\beta}_{MCO}$ tiene una distribución normal multivariante exacta con media β y matriz de varianzas y covarianzas $\sigma^2(X'X)^{-1}$. La prueba se basa sobre una propiedad de la distribución normal multivariante, por la cual una transformación lineal de un vector aleatorio distribuido según una normal multivariante sigue también una distribución normal multivariante: sea $Z \sim N(\mu, \Sigma)$, entonces una cualquiera transformación lineal de Z, $AZ + b$, se distribuye como: $(AZ + b) \sim N(A\mu + b, A\Sigma A')$. Se observa que $\hat{\beta}_{MCO}$ es una transformación lineal del vector ε, dado que $\hat{\beta}_{MCO} = \beta + (X'X)^{-1}X'\varepsilon$; siendo $\varepsilon \sim N(0, \sigma^2 I)$, podemos expresar $\hat{\beta}_{MCO} = A\varepsilon + b$, donde en ese caso $A = (X'X)^{-1}X'$, $b = \beta$, $\mu = 0$ y $\Sigma = \sigma^2 I$:

$$\begin{aligned}
\hat{\beta}_{MCO} &= (\beta + (X'X)^{-1}X'\varepsilon) \sim N((X'X)^{-1}X' \cdot 0 + \beta, (X'X)^{-1}X'\sigma^2 I[(X'X)^{-1}X']') \\
\hat{\beta}_{MCO} &\sim N(0 + \beta, (X'X)^{-1}X'\sigma^2 IX(X'X)^{-1}) \\
\hat{\beta}_{MCO} &\sim N(\beta, \sigma^2(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{=I_k}) \\
\hat{\beta}_{MCO} &\sim N(\beta, \underset{(k\times k)}{\sigma^2(X'X)^{-1}})
\end{aligned}$$

El elemento j del vector $\hat{\beta}_{MCO}$, $\hat{\beta}_j$, donde $1 \leq j \leq k$, corresponde al estimador de MCO del parametro β_j y tiene una distribución normal univariante con media β_j y varianza correspondiente al elemento (j, j) en la diagonal de la matriz de varianza y covarianza $\sigma^2(X'X)^{-1}$:

$$\hat{\beta}_j \sim N(\beta_j, [\sigma^2(X'X)^{-1}]_{(j,j)})$$

El supuesto H6 de normalidad de los errores implica también que la variable dependiente y tenga una distribución normal multivariante, siendo una transformación lineal de ε. Dado que $E(y) = X\beta$, la varianza de y coincide con la de ε, $var(y) = var(\varepsilon)$, y:

$$y = (X\beta + \varepsilon) \sim N(X\beta, \sigma^2 I_N)$$

Igualmente, es posible establecer la distribución de la *predicción* de y, $\hat{y} = X\hat{\beta}_{MCO}$, definida como el valor estimado de y cuando se estima β con $\hat{\beta}_{MCO}$. $X\hat{\beta}_{MCO}$ es una transformación lineal de $\hat{\beta}_{MCO}$, entonces por la propiedad de la transformación lineal de una distribución normal multivariante, si $\hat{\beta}_{MCO} \sim N(\beta, \sigma^2(X'X)^{-1})$:

$$\begin{aligned}
\hat{y} &= X\hat{\beta}_{MCO} \sim N(X\beta, X[\sigma^2(X'X)^{-1}]X') \\
\hat{y} &\sim N(X\beta, \sigma^2 X(X'X)^{-1}X')
\end{aligned}$$