DOCTORADO EN ECONOMÍA APLICADA CON MENCIÓN EN ESTUDIOS REGIONALES

EXAMEN DE CALIFICACIÓN
ECONOMETRÍA/ECONOMETRÍA ESPACIAL
Pauta de corrección

Departamento de Economía, UCN
Marzo, 2022

1. Suponga que usted está interesado en obtener un estimador de la variable y_i a partir de x_i, donde x_i es un vector de variables correlacionadas con y_i.
   a. Sea $m(\mathrm{x}_i)$ una función no restringida de x_i. Demuestre que la Función de Esperanza Condicional $E(y_i \mid \mathrm{x}_i)$, es el mejor estimador no restringido de y_i. (10 puntos)

      Solución:
      El mejor estimador es el que minimiza el Error Cuadrático Medio, por lo tanto, lo que se pide demostrar es: $E(y_i|\mathrm{x}_i) = \underset{m(\mathrm{x}_i)}{\operatorname{argmin}} E\left[(y_i - m(\mathrm{x}_i))^2 |\mathrm{x}_i\right]$.
      Resolviendo el problema de optimización:

$$\frac{\partial E\left[(y_i - m(\mathrm{x}_i))^2 |\mathrm{x}_i\right]}{\partial m(\mathrm{x}_i)} = E\left[-2(y_i - m(\mathrm{x}_i))|\mathrm{x}_i\right] = 0$$

$$E\left[(y_i - m(\mathrm{x}_i))|\mathrm{x}_i\right] = 0$$

$$E(y_i|\mathrm{x}_i) - E(m(\mathrm{x}_i)|\mathrm{x}_i) = 0$$

$$E(y_i|\mathrm{x}_i) - m(\mathrm{x}_i) = 0$$

$$E(y_i|\mathrm{x}_i) = m(\mathrm{x}_i)$$

   b. Suponga ahora que $E(y_i \mid \mathrm{x}_i)$ es una función restringida de x_i, tal que $E(y_i \mid \mathrm{x}_i)$ es lineal en un vector de parámetros b: $E(y_i \mid \mathrm{x}_i) = \mathrm{x}_i'\mathrm{b}$. Donde b es de dimensión $k \times 1$. Demuestre que el estimador de Mínimos Cuadrados Ordinarios es el mejor predictor de b dada una muestra de tamaño $n >> k$. (10 puntos)

Solución:

Se nos pide demostrar que si $E\left(y_i \mid \mathrm{x}_i\right)$ es lineal, entonces el estimador de Mínimos Cuadrados Ordinarios será el mejor estimador del vector de parámetros b para una muestra dada. Planteamos el problema en términos poblacionales: $\beta = \underset{\mathrm{b}}{\operatorname{argmin}} E\left[\left(y_i - \mathrm{x}_i'\mathrm{b}\right)^2 \left|\mathrm{x}_\mathrm{i}\right.\right]$.

Resolviendo el problema de optimización:

$$\frac{\partial E\left[\left(y_i - \mathrm{x}_i'\mathrm{b}\right)^2 \left|\mathrm{x}_\mathrm{i}\right.\right]}{\partial b} = E\left[-2\mathrm{x}_i\left(y_i - \mathrm{x}_i'\mathrm{b}\right) \mid \mathrm{x}_i\right] = 0$$

$$\Rightarrow \beta = \left[E\left(\mathrm{x}_i\mathrm{x}_i'\right)\right]^{-1} E\left(\mathrm{x}_i y_i\right)$$

Pero el análogo de la expresión anterior para una muestra de tamaño n es justamente el estimador de MCO:

$$\hat{\beta} = \left(\frac{1}{n}\sum_{i=1}^{n}\mathrm{x}_\mathrm{i}\mathrm{x}_\mathrm{i}'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n}\mathrm{x}_\mathrm{i}y_i = \left(\sum_{i=1}^{n}\mathrm{x}_\mathrm{i}\mathrm{x}_\mathrm{i}'\right)^{-1}\sum_{i=1}^{n}\mathrm{x}_\mathrm{i}y_i$$

Nota: Ya que el Error Cuadrático Medio es igual al sesgo al cuadrado más la varianza, como solución alternativa vale también demostrar que el estimador de MCO es insesgado y de mínima varianza entre los lineales insesgados (Teorema Gauss-Markov). Sin embargo, este es un camino más largo para arribar a la solución.

c. Suponga ahora que $E\left(y_i \mid \mathrm{x}_i\right)$ no es lineal. ¿Cuál es la mejor aproximación lineal de $E\left(y_i \mid \mathrm{x}_i\right)$ dada una muestra de tamaño $n >> k$? (5 puntos)

Solución:

Si $E\left(y_i \mid \mathrm{x}_i\right)$ no es lineal, entonces el mejor estimador lineal de $E\left(y_i \mid \mathrm{x}_i\right)$ es $y_i = \mathrm{x}_\mathrm{i}'\hat{\beta}$, donde $\hat{\beta}$ es el estimador de MCO, puesto que como ya se demostró en el punto anterior, $\hat{\beta}$ es el mejor estimador lineal de b.

2. Usted sabe que el modelo correcto es $y_i = \mathrm{x}_{i'}\beta + u_i$, tal que $E(\mathrm{x}_i u_i) = 0$. Sin embargo, las variables explicativas están medidas con un error aleatorio ϵ_i, de manera que usted solamente observa: $\mathrm{x}_i^* = \mathrm{x}_i + \mathrm{e}_i$, donde e_i es el vector de errores de medición. Usted sabe además que: $E(\mathrm{x}_i\mathrm{e}_{i'}) = 0$, $E(u_i\mathrm{e}_i) = 0$ y $E\left(\mathrm{e}_i\mathrm{e}_i'\right) = \Sigma_\epsilon$.

   a. ¿Cuál sería el estimador de MCO en función de e_i, si usted estima el modelo observado: $y_i = \mathrm{x}_i^{*\prime}\beta + v_i$? Donde v_i es la perturbación aleatoria del modelo observado. (10 puntos)

Solución:

Se trata de un problema típico de endogeneidad por error de medición. El estimador de MCO en este caso será el siguiente:

$$\hat{\beta} = \left(\sum_i \mathrm{x}_i^* \mathrm{x}_i^{*}{}' \right)^{-1} \sum_i \mathrm{x}_i^* y_i$$

Encontramos ahora una expresión para el modelo verdadero en función de $\mathrm{e_i}$: $y_i = \mathrm{x}_{i'}\beta + u_i = \left(\mathrm{x}_i^* - \mathrm{e}_i\right)' \beta + u_i = \mathrm{x}_i^{*\prime}\beta - \mathrm{e}_{i'}\beta + u_i$. Ahora sustituimos en el estimador de MCO:

$$\hat{\beta} = \left(\sum_i \mathrm{x}_i^* \mathrm{x}_i^{*}{}' \right)^{-1} \sum_i \mathrm{x}_i^* y_i$$

$$\hat{\beta} = \left(\sum_i \mathrm{x}_i^* \mathrm{x}_i^{*}{}' \right)^{-1} \sum_i \mathrm{x}_i^* \left(\mathrm{x}_i^{*\prime}\beta - \mathrm{e}_{i'}\beta + u_i \right)$$

$$\hat{\beta} = \left(\sum_i \mathrm{x}_i^* \mathrm{x}_i^{*}{}' \right)^{-1} \sum_i \mathrm{x}_i^* \mathrm{x}_i^{*\prime}\beta + \left(\sum_i \mathrm{x}_i^* \mathrm{x}_i^{*}{}' \right)^{-1} \sum_i \mathrm{x}_i^* \left(-\mathrm{e}_{i'}\beta + u_i \right)$$

$$\hat{\beta} = \beta + \left(\sum_i \mathrm{x}_i^* \mathrm{x}_i^{*}{}' \right)^{-1} \sum_i \mathrm{x}_i^* \left(-\mathrm{e}_{i'}\beta + u_i \right)$$

$$\hat{\beta} = \beta + \left(\sum_i \mathrm{x}_i^* \mathrm{x}_i^{*}{}' \right)^{-1} \sum_i \left(\mathrm{x}_i + \mathrm{e}_i \right)\left(-\mathrm{e}_{i'}\beta + u_i \right)$$

b. ¿Cuál es el sesgo asintótico del estimador de MCO encontrado en el punto anterior? (10 puntos)

Solución:

Expresamos el estimador de MCO como función de n:

$$\hat{\beta} = \beta + \left(\frac{\sum_i \mathrm{x}_i^* \mathrm{x}_i^{*}{}'}{n} \right)^{-1} \frac{\sum_i \left(\mathrm{x}_i + \mathrm{e}_i \right)\left(-\mathrm{e}_{i'}\beta + u_i \right)}{n}$$

Por Ley Débil de Grandes Números las medias muestrales convergen en probabilidad a las medias poblacionales:

$$\hat{\beta} \xrightarrow{p} \beta + E\left(\mathrm{x}_i^* \mathrm{x}_i^{*}{}' \right)^{-1} E\left[\left(\mathrm{x}_i + \mathrm{e}_i \right)\left(-\mathrm{e}_{i'}\beta + u_i \right) \right]$$

Simplificando el último término:

$$E\left[\left(\mathrm{x}_i+\mathrm{e}_i\right)\left(-\mathrm{e}_i'\beta+u_i\right)\right]=$$
$$-E\left(\mathrm{x}_i\mathrm{e}_i'\right)\beta+E\left(\mathrm{x}_iu_i\right)-E\left(\mathrm{e}_i\mathrm{e}_i'\right)\beta+E\left(\mathrm{e}_iu_i\right)$$
$$E\left[\left(\mathrm{x}_i+\mathrm{e}_i\right)\left(-\mathrm{e}_i'\beta+u_i\right)\right]=-E\left(\mathrm{e}_i\mathrm{e}_i'\right)\beta$$

Sustituyendo:

$$\hat{\beta}\xrightarrow{p}\beta+E\left(\mathrm{x}_i^*\mathrm{x}_i^*{}'\right)^{-1}\left[-E\left(\mathrm{e}_i\mathrm{e}_i'\right)\beta\right]$$
$$\hat{\beta}\xrightarrow{p}\beta+E\left(\mathrm{x}_i^*\mathrm{x}_i^*{}'\right)^{-1}\left(-\Sigma_\varepsilon\beta\right)$$

Por lo tanto, el sesgo asintótico es igual a: $E\left(\mathrm{x}_i^*\mathrm{x}_i^{*'}\right)^{-1}\left(-\Sigma_\varepsilon\beta\right)$.

3. De acuerdo con el siguiente modelo,

$$y = Z\beta + u$$
$$u = \rho W u + \varepsilon$$

Donde $\varepsilon \sim N(0, \sigma^2 I_n)$ y $|\rho| < 1$. La matriz W es no-estocástica.

a) Demuestre que la interpretación de los coeficientes es igual en comparación a un modelo lineal no espacial.

<u>Solución:</u>

Para demostrar esto, es importante notar que el modelo es un Spatial Error Model (SEM). Podemos expresar el modelo de la siguiente forma:

$$y = Z\beta + (I - \rho W)^{-1}\varepsilon$$

Donde

$$u = (I - \rho W)^{-1}\varepsilon$$

Obtenemos el valor esperando de la variable dependiente:

$$E[Y|Z] = Z\beta + (I - \rho W)^{-1}E(\varepsilon|Z)$$

Dado que $E(\varepsilon|Z) = 0$, es el mismo valor esperado en comparación a un modelo lineal estándar.

Para interpretar los coeficientes de regresión estadísticamente, obtenemos la matriz de derivadas parciales para una variable exógena k, para corroborar que su estructura funcional es igual a un modelo lineal estándar. Considerando n observaciones:

$$\begin{bmatrix} \frac{\partial y_1}{\partial x_{1k}} & \cdots & \frac{\partial y_1}{\partial x_{nk}} \\ \vdots & \ddots & \vdots \\ \frac{\delta y_n}{\delta x_{1k}} & \cdots & \frac{\partial y_n}{\partial x_{nk}} \end{bmatrix} = \begin{bmatrix} \beta_k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \beta_k \end{bmatrix} = \beta_k \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = \beta_k I_n$$

Donde las derivadas cruzadas de la matriz son iguales a cero, y solo los términos de la diagonal son distintos de cero, evidenciando la no existencia de efectos spillovers (derrame). Es decir, las observaciones son independientes y los cambios que ocurren en alguna observación no afectan al resto de las observaciones.

La única diferencia entre el modelo SEM y un modelo lineal esta en el error del modelo SEM, ya que depende de una matriz W. Sin embargo, debido a los valores esperados son iguales la interpretación de los coeficientes es igual.

b) Demuestre matemáticamente porqué la estimación de este modelo a través de OLS pierde sus propiedades óptimas.

Solución:

Podemos expresar el modelo SEM de la siguiente forma, considerado que $(I - \rho W)u = \varepsilon$ y $y - Z\beta = u$,

$$\begin{aligned} (I - \rho W)u &= \varepsilon \\ (I - \rho W)(y - Z\beta) &= \varepsilon \\ y - Z\beta - \rho W y + W Z \rho \beta &= \varepsilon \end{aligned}$$

Reorganizando la expresión y definiendo $\gamma = \rho\beta$

$$y = Z\beta + \rho W y - WZ\,\gamma + \varepsilon$$

Desde aquí, debemos demostrar que el término Wy (el rezago espacial) está correlacionado con el término de error, lo que produce endogeneidad. Para ello, re-escribimos la expresión anterior como:

$$y = (I - \rho W)^{-1}(Z\beta - WZ\,\gamma) + (I - \rho W)^{-1}\varepsilon$$

Obtenemos la covarianza entre el término Wy y el término de error:

$$\begin{aligned} E[(Wy)\varepsilon'] &= E[W((I - \rho W)^{-1}(Z\beta - WZ\,\gamma) + (I - \rho W)^{-1}\varepsilon)\varepsilon'] \\ E[(Wy)\varepsilon'] &= W(I - \rho W)^{-1}(Z\beta - WZ\,\gamma)E[\varepsilon'] + W(I - \rho W)^{-1}E[\varepsilon\varepsilon'] \end{aligned}$$

Ya que $E[\varepsilon'] = 0$

$$
\begin{aligned}
E[(Wy)\varepsilon'] &= W(I-\rho W)^{-1}E[\varepsilon\varepsilon'] \\
E[(Wy)\varepsilon'] &= E[\varepsilon\varepsilon']W(I-\rho W)^{-1} \\
E[(Wy)\varepsilon'] &= \sigma_\varepsilon^2 W(I-\rho W)^{-1}I_n \neq 0
\end{aligned}
$$

Queda demostrado que el error es endógeno, ya que está correlacionado el rezago espacial del modelo. De esta forma, OLS pierde sus propiedades óptimas.

c) ¿Por qué no es posible usar variables instrumentales para obtener estimadores consistentes?

Solución:

Según Kelejian and Prucha (1998) no es posible aplicar variables instrumentales porque no es posible encontrar variables instrumentales para Wy, que sean independientes de Z y WZ.

d) Indique los pasos para obtener los parámetros del modelo a través de Máxima Verosimilitud. ¿Por qué la función de log-verosimilitud obtenida no puede ser maximizada de forma analítica? ¿Cómo es posible maximizar la expresión?

Solución:

Paso 1: Asumiendo que u está normalmente distribuido, se construye la función de verosimilitud.

Paso 2: Se escribe el logaritmo de la verosimilitud. De esta forma, obtenemos la función de soporte.

Paso 3: Maximizamos la función de soporte para obtener los parámetros desconocidos. Estas son las condiciones de primer orden.

Paso 4: Comprobamos que los parámetros obtenidos en el paso 3 son máximos.

La función de soporte (la función de log-verosimilitud) no es posible maximizar analíticamente, debido a que es una función altamente no lineal. Por tanto, es necesario obtener el valor de los parámetros de forma numérica, usando una función de log-verosimilitud concentrada, con respecto a β y σ^2, para obtener una función de log-verosimilitud en función de ρ.