Econometría de Datos de Panel

Maestrías en Economía y Econometría

Examen Final
Primer Trimestre 2024

1 Propiedades de Muestra Finita en Paneles No Balanceados

Considere el siguiente modelo:

$$y_{jt}^* = \beta x_{jt} + c_j + u_{jt}, \quad j = 1, 2, ...N; \; t = 1, 2, ..., T. \tag{1}$$

$$s_{jt}^* = \gamma_1 w_{jt} + \gamma_2 z_{jt} + \alpha_j + \epsilon_{jt}$$

$$s_{jt} = 1, \text{ sí } s_{jt}^* > 0, \quad s_{jt} = 0, \text{ en cualquier otro caso}$$

$$y_{jt} = y_{jt}^* \times s_{jt}$$

Las dos variables de la ecuación de selección (w y z) son independientes y normalmente distribuídas con media cero y varianza uno. La única variable de la ecuación de interés es $x_{jt} = w_{jt}$, por lo tanto tenemos una variable que está excluída de la ecuación de interés. El error idiosincrático de la ecuación de selección, ϵ_{jt}, sigue una distribución distribución normal con media cero y varianza uno. El error idiosincrático de la ecuación de interés (8) se define como: $u_{jt} = 0.6\epsilon_{jt} + 0.8\psi_1$, con ψ_1 siendo una variable independiente y normal estándar. Con esta especificación la correlación entre los errores idiosincráticos de las ecuaciones de interés y de selección es 0.6. Además defina tres variables aleatotias: ψ_2, ψ_3 y ψ_4. Estas variables son específicas de corte transversal. Son todas independientes y normales estándar. Con esta especificación general defina tres modelos:

Modelo A: este modelo asume que los dos términos de efectos no observables, c_j y α_j están correlacionados. Más específicamente:

$$\alpha_j = \psi_2 + \psi_4 \tag{2}$$

$$c_j = \psi_3 + \psi_4 \tag{3}$$

Modelo B: este modelo asume que las variables observables y no observables están correlacionadas, pero que no hay correlación entre los efectos no observables entre ecuaciones.

$$\alpha_j = \psi_2 + \frac{z_{j1} + z_{j2}}{2} \tag{4}$$

$$c_j = \psi_3 + \frac{x_{j1} + x_{j2}}{2} \tag{5}$$

Modelo C: este es el modelo más general porque permite ambos tipos de correlación entre los dos componentes no observables y entre estos componentes y las variables observables.

$$\alpha_j = \psi_2 + \frac{z_{j1} + z_{j2}}{2} + \psi_4 \tag{6}$$

$$c_j = \psi_3 + \frac{x_{j1} + x_{j2}}{2} + \psi_4 \tag{7}$$

Todos los parámetros son iguales a uno ($\beta = \gamma_1 = \gamma_2 = 1$). $N = 20, 40, 100$ y $T = 2, 10$.

1. Caso 1: Para cada modelo y combinación de $N = 20, 40, 100$ y $T = 2$ realice un experimento de Monte Carlo de $S = 1000$ simulaciones y reporte el sesgo medio, el sesgo mediano, el error estándar, el RMSE y la desviación media absoluta de la estimación de Wooldridge para los tres parámetros del modelo (β, γ_1, y γ_2). Explique sus resultados.

2. Caso 2: Para cada modelo y combinación de $N = 20, 40, 100$ y $T = 10$ considere lo siguiente.

   (a) Tome la primera generación de los datos ($S = 1$) como si fuera una muestra obtenida de la realidad.

   (b) Estime el modelo por Wooldridge y guarde los coeficientes estimados (llámelos $\beta^{(b)}$, $\gamma_1^{(b)}$, y $\gamma_2^{(b)}$) como así también los residuos de la ecuación de selección, $\hat{\epsilon}_{jt}^{(b)}$.

   (c) Realice un procedimiento de bootstrapping para construir intervalos de 95% de confiabilidad para los tres estimadores de la siguiente manera: (i) utilice un procedimiento de muestreo aleatorio simple con reemplazo para obtener una nueva muestra de errores de la ecuación de selección (Nota: observe que debe generar para cada j una nueva muestra de residuos $\hat{\epsilon}_{jt}^{(b)}$ de dimensión T). Con estos nuevos residuos, con $\gamma_1^{(b)}$, y $\gamma_2^{(b)}$ y con los valores originales de w, z (es decir, con los valores generados para $S = 1$) y α_j construya una nueva variable dependiente del modelo de selección. (ii) Usando esta nueva variable dependiente del modelo de selección, los valores de $\beta^{(b)}$ y x_{jt} y c_j construya una nueva variable dependiente de la ecuación de interés. (iii) Con las nuevas variables dependientes de ambas ecuaciones construídas, estime por Wooldridge los tres parámetros del modelo. (iv) Vuelva al paso (i) y repita el procedimiento. Repita (iv) para construir $B = 1000$ muestras de bootstrapping y en cada caso guarde los coeficientes estimados por Wooldridge de β, γ_1, y γ_2. Reporte, para $B = 1000$, intervalos de bootstrapping de 95% de confiabilidad. Compare estos resultados con intervalos de 95% de significatividad construídos a partir de la matriz de varianzas y covarianzas asintótica para el modelo estimado con los datos de $S = 1$. Qué conclusiones puede sacar?.

2 Propiedades de Muestra Finita en Paneles Dinámicos

Recientemente, ha habido un renovado interés de la literatura macroeconómica en examinar el crecimiento en el largo plazo utilizando modelos dinámicos con datos de panel para países (vea Mankiw, Romer y Weil, 1992; Levine y Renelt, 1992 entre otros). También la literatura de economía laboral ha utilizado los modelos dinámicos, por ejemplo, para el análisis de la relación existente entre salarios y desempleo, la denominada "wage curve" por Blanchflower y Oswald (1994) (vea Blanchard y Katz, 1997; Galiani, 1999).

En general la dimensión de los macropaneles para analizar este tipo de problemas está caracterizada por un valor de T entre 10 y 30 y un valor de N entre 40 y 50.

En términos de los modelos tradicionales de datos de panel, sabemos que LSDV es sesgado e inconsistente para T fijo y N tendiendo a infinito; mientras que los estimadores que utilizan procedimientos de IV o GMM son consistentes cuando N tiende a infinito.

En este trabajo usted analizará como estimar y hacer inferencia en paneles que no tienen una dimensión muy pequeña en T y no tienen una dimensión muy grande en N. Por ejemplo, un valor de $T = 30$ es lo suficientemente grande como para ignorar el sesgo de LSDV? ó un valor de $N = 50$ es lo suficientemente grande para que los métodos de IV o GMM den estimadores consistentes?

Para responder este tipo de interrogantes, considere el siguiente modelo:

$$y_{jt} = \alpha y_{jt-1} + \beta x_{jt} + c_j + u_{jt}, \quad j = 1, 2, ...N; \; t = 1, 2, ..., T + 10. \tag{8}$$

con $c_j \sim IN(0, 1)$, $u_{jt} \sim IN(0,\ 1)$, $y_{j0} = 0$ y las primeras 10 observaciones de series temporales se descartan, es decir que el tamaño muestral es NT. El regresor adicional, x_{jt}, es una variable estríctamente exógena generada de la siguiente manera: $x_{jt} = 0.8x_{jt-1} + v_{jt}$, donde $v_{jt} \sim N(0,\ 0.9)$

1. Caso 1: $\beta = 0$, $T = 10$, $N = 30$, $\alpha = 0.5$. Realice un experimento de Monte Carlo con 1000 simulaciones. Reporte media, desvío estándar y RMSE de la estimación de α usando: LSDV, AB-GMM1, AB-GMM2, BB-GMM1, AH y Kiviet. Reporte el tamaño del test para $\alpha = 0.5$ en todos los casos. Comente los resultados obtenidos y su conclusión de qué estimador debiera utilizarse en la práctica. (Nota: AH es el estimador de Anderson-Hsiao que utiliza los niveles rezagados como instrumento para la primera diferencia; Kiviet es

el estimador de LSDV corregido por el sesgo asintótico. En este último caso utilice AH como estimación para construir la matriz C).

2. Repita el punto anterior con $N = 50$. Compare los resultados en ambos puntos.

3. Ahora suponga que $\beta = 0$, $T = 20$, $N = 30$, $\alpha = 0.8$. Repita el ejercicio anterior.

4. Ahora suponga que $\beta = 0$, $T = 20$, $N = 30$, $\alpha = 0.92$. Repita el ejercicio anterior prestando particular atención al comentario de BB acerca de los instrumentos débiles. Qué puede comentar al respecto.

5. Ahora suponga que $\beta = 0$, $T = 30$, $N = 50$, $\alpha = 0.5$. Repita el ejercicio anterior prestando particular atención al sesgo del estimador LSDV. Es $T = 30$ suficiente como para ignorar el sesgo de LSDV?

6. Repita el punto 1 con: $\beta = 1$, $T = 7$, $N = 100$, $\alpha = 0.8$. Compare sus resultados con los obtenidos por AB, Tabla 1, página 284. Cuáles son sus conclusiones? (Nota: ahora tiene que reportar también sus resultados para la estimación de β.)

7. Repita el punto 1 con: $\beta = 0$, $T = 4$, $N = 100$, $\alpha = 0.8$. Compare sus resultados con los obtenidos por BB, Tabla 2(a), página 131.

8. Comparando los resultados de los puntos 3, 6 y 7, puede sacar alguna conclusión acerca del comportamiento de los estimadores para las dimensiones del panel en 3 versus la dimensión en 6 y 7?

3 Elasticidad Precio de Demanda de los Cigarrillos

En este ejercicio se le pide estimar una serie de modelos de panel para tratar de calcular la elasticidad precio de la demanda de los cigarrillos. Los datos, cigs.dta, tienen las siguientes columnas:

- sales: ventas de cigarrillos en paquetes per cápita
- pimin: precio mínimo en estados cercanos por paquete de cigarrillos
- ndi: ingreso disponible per cápita
- cpi: índice de precios al consumidor (1983=100)
- pop16: población mayor a los 16 años
- pop: población total
- price: precio por paquete de cigarrillos
- year: el año
- state: estado.

1. La elasticidad precio de demanda, E_d es el cambio porcentual en las ventas de cigarrillos dividido por el cambio porcentual en el precio de los cigarrilos. Sea S_t las ventas de cigarrillos en el período t y p_t el precio de los cigarrillos en el período t, entonces la elasticidad precio de demanda es:

$$E_d = \frac{\Delta S_t / S_t}{\Delta p_t / p_t}$$

   Para estimar la elasticidad en un modelo lineal aplique logaritmos a las ventas y el precio y estime por POLS

$$\log\left(S_{i,t}\right) = \gamma + \beta log\left(p_{i,t}\right) + \alpha_1 ndi_{i,t} + \alpha_2 cpi_{i,t} + \alpha_3 pop16_{i,t} + \epsilon_{i,t}$$

   Reporte la estimación de la elasticidad precio de demanda? Qué puede estar potencialmente mal en esta estimación?

2. Una característica importante puede ser la heterogeneidad entre estados. Sin embargo no sabemos si queremos modelar esta heterogeneidad como un efecto fijo o aleatorio. Realice un contraste de hipótesis para determinar que modelización usar. Interprete el resultado del contraste a niveles de signicación usuales.

3. Empezamos a creer que no solo la heterogeneidad es importante sino también la dinámica así que decidimos estimar el siguiente modelo:

$$\log(S_{i,t}) = \gamma_i + \tau \log(S_{i,t-1}) + \beta \log(p_{i,t}) + \alpha_1 ndi_{i,t} + \alpha_2 cpi_{i,t} + \alpha_3 pop16_{i,t} + \epsilon_{i,t}$$

   Aplique la transformación de diferencias finitas al modelo del punto anterior. Sea $\Delta S_{i,t} = \log(S_{i,t}) - \log(S_{i,t-1})$, la variable dependiente de una estimación de mínimos cuadrados en dos etapas usando el procedimiento de Anderson-Hsiao. Reporte sus resultados y cualquier diferencia con sus estimaciones anteriores.

4. Re-estime el modelo usando el estimador de Arellano-Bond. Encuentra alguna diferencia con sus estimaciones anteriores? Interprete sus resultados.

Referencias

Anderson, T.W: and Cheng Hsiao (1981) (AH), "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association* 76, 598-606.

Arellano M. and Stephen Bond (1991) (AB), "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *The Review of Economic Studies* 58, 277-297.

Blanchard O. and Lawrence Katz (1997), "What We Know and Do Not Know About the Natural Rate of Unemployment," *Journal of Economic Perspectives* 11, 51-72.

Blanchflower, D. and A. Oswald (1996), The wage curve, The MIT Press, Cambridge, Massachusetts.

Blundell, R. and Stephen Bond (1998) (BB), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics* 87, 115-143.

Galiani, S., 1999, "Wage determination in Argentina: An econometric analysis with methodology discussion," ITDT WP 218.

Kiviet Jan (1995), "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models," *Journal of Econometrics* 68, 53-78.

Levine, R. and D. Renelt (1992), "A Sensitivity Analysis of Cross-Country Growth Regressions," *American Economic Review*, September 1992, 942-63.

Mankiw, N.G., D. Romer and D. Weil (1992) "A Contribution to the Empirics of Economic Growth," *Quarterly Journal of Economics*, May 1992, 407-38.

Instrucciones para la entrega:

- El trabajo es individual.
- Use los últimos 4 números de su documento de identidad como seed en el código para que los resultados sean replicables.
- El trabajo se puede hacer en Stata, R, Matlab o similar.
- Hay que entregar un archivo .pdf (.doc o similar) con un reporte con todas las respuestas a las preguntas del trabajo y adjuntar el código (.do, .r, .m, etc.) o entregar el .log file aparte.
- En el código deberán hacer comentarios breves para que se entienda el procedimiento.
- La entrega se realizará vía Campus Virtual hasta el día lunes 24 de junio de 2024 a las 11.59pm.