

Universidad Torcuato Di Tella
Maestría en Economía y Maestría en Econometría
Econometría de Datos de Panel
Ejercicio Número 2.
9 de mayo de 2023

Propiedades de Muestra Finita en Paneles No Balanceados

Considere el siguiente modelo:

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

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

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

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

Las dos variables de la ecuación de selección (w y z) son independientes y normalmente distribuidas 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 excluida de la ecuación de interés. El error idiosincrático de la ecuación de selección, ϵ_{jt} , sigue una distribución normal con media cero y varianza uno. El error idiosincrático de la ecuación de interés (1) 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 aleatorias: 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:

$$\alpha_j = \psi_2 + \psi_4 \quad (2)$$

$$c_j = \psi_3 + \psi_4 \quad (3)$$

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.

$$\alpha_j = \psi_2 + \frac{z_{j1} + z_{j2}}{2} \quad (4)$$

$$c_j = \psi_3 + \frac{x_{j1} + x_{j2}}{2} \quad (5)$$

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.

$$\alpha_j = \psi_2 + \frac{z_{j1} + z_{j2}}{2} + \psi_4 \quad (6)$$

$$c_j = \psi_3 + \frac{x_{j1} + x_{j2}}{2} + \psi_4 \quad (7)$$

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?.

Modelos de Panel no Lineales

El archivo KEANE.DTA contiene un subconjunto de los datos utilizados por Keane y Wolpin¹ en formato Stata. Estos datos tienen información sobre empleo e historia escolar para una muestra de hombres para los años 1981-1987. El archivo KEANE.DES contiene una detallada descripción de los datos. Los dos archivos se adjuntan con el examen.

Para responder esta pregunta considere solo los hombres negros.

- (a) Use Pooled Probit para estimar el modelo:

$$P(\text{employ}_{jt} = 1 | \text{employ}_{j,t-1}) = \Phi(\delta_0 + \rho \text{employ}_{j,t-1}).$$

¹Keane, M.P. and K.I. Wolpin (1997), "The Career Decisions of Young Men, *Journal of Political Economy* 105, 473-522.

Qué supuestos son necesarios para asegurar que los desvíos estándar y estadísticos t de pooled probit sean asintóticamente válidos?

- (b) Estime $P(employ_{jt} = 1 | employ_{j,t-1} = 1)$ y $P(employ_{jt} = 1 | employ_{j,t-1} = 0)$. Explique cómo obtendría errores estándar para estas estimaciones.
- (c) Adicione al modelo estimado en 1. el conjunto completo de variables binarias temporales y estime las probabilidades pedidas en (b). para 1987. Hay diferencias importantes con las estimaciones de (b).?
- (d) Ahora estime un modelo de efectos no observables dinámico. En particular, adicione $employ_{j,81}$ como una variable explicativa adicional y use el modelo Probit de efectos aleatorios. Use el conjunto completo de variables binarias temporales.
- (e) Hay evidencia de dependencia del estado anterior, condicional en c_j ? Explique.
- (f) Promedie las probabilidades estimadas a lo largo de $employ_{j,81}$ para obtener el efecto parcial promedio para 1987. Compare estas estimaciones con aquellas en (c).

Instrucciones para la entrega:

- El trabajo es individual.
- Use los últimos 4 números de su documento como seed en el archivo .do para que los resultados sean replicables.
- El trabajo se puede hacer en Stata, R, Python o similar.
- Hay que entregar un archivo .pdf con el reporte y adjuntar el código (.do, .r, etc.) o entregar el .log file aparte.
- Los archivos deberán tener el apellido de quien entregue y la seed utilizada, es decir, si Caravello usó el seed 1668 los archivos se llamarán Caravello1668.pdf, Caravello1668.do, Caravello1668.log, Caravello1668.r, etc.
- En el código deberán hacer comentarios breves para que se entienda el procedimiento.
- La entrega se realizará vía Campus Virtual hasta el día lunes 26 de junio de 2023 a las 11.59pm.