Estimación bajo exogeneidad secuencial

Gustavo A. García

ggarci24@eafit.edu.co

Econometría avanzada II

PhD/Maestría en Economía

Universidad EAFIT

Link slides en formato html

Link slides en formato PDF

En este tema

- Motivación
- Marco general
- Estimación bajo exogeneidad secuencial
- Modelos con variable dependiente rezagada
- Ejercicio aplicado en R: modelos dinámicos con datos de panel

Lecturas

- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data*. 2a edición. MA: MIT Press. Sección 11.6

Motivación

- Los métodos de estimación usualmente asumen que las variables explicatorias son estrictamente exógenas (condicional sobre un efectos inobservables en el caso de efectos fijos)

- Generalmente, efectos aleatorios y fijos son inconsistentes si una variable explicatoria en algún periodo de tiempo es correlacionado con u_{it}

- Se debe, entonces, tener una forma general de obtener consistentes estimadores cuando $N \longrightarrow \infty$ con T es fijo cuando las variables explicatorias no son estrictamente exógenas

Marco general

El modelo de interés es

$$y_{it} = \mathbf{x}_{it}\beta + c_i + u_{it}, t = 1, 2, \ldots, T$$

c_i y $\mathbf{x}_{it}$ pueden estar correlacionadas y adicionalmente se permite que u_{it} a estar correlacionada con valores *futuros* de las variables explicatorias $(\mathbf{x}_{i,t+1}, \mathbf{x}_{i,t+2}, \ldots, \mathbf{x}_{iT})$

u_{it} y $\mathbf{x}_{i,t+1}$ deben estar correlacionadas ya que $x_{i,t+1} = y_{it}$

No obstante, existen muchos modelos, incluyendo el modelo AR(1), para lo cual es razonable asumir que u_{it} esta correlacionado con actuales y pasados valores de $\mathbf{x}_{it}$

Para tratar esta correlación, Chamberlain (1992) introduce la \restricción de momentos secuenciales:

$$E(u_{it}|\mathbf{x}_{it}, \mathbf{x}_{i,t-1}, \ldots, \mathbf{x}_{i1}, c_i) = 0, t = 1, 2, \ldots, T$$

Cuando el anterior supuesto se mantiene, decimos que las $\mathbf{x}_{it}$ son secuencialmente exógenas condicional a los efectos inobservables

Marco general

El anterior supuesto es equivalente a

$$E(y_{it}|\mathbf{x}_{it}, \mathbf{x}_{i,t-1}, \ldots, \mathbf{x}_{i1}, c_i) = E(y_{it}|\mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\boldsymbol{\beta} + c_i$$

es claro entonces lo que implica la exogeneidad secuencial en las variables explicatorias: después de que $\mathbf{x}_{it}$ y c_i han sido controladas, valores pasados de $\mathbf{x}_{it}$ no afectan el valor esperado de y_{it}

Esta condición es más natural que el supuesto de exogeneidad estricta, la cual requiere también condicionar sobre valores futuros de $\mathbf{x}_{it}$

Marco general

Ejemplo: Modelo estático con *feedback*

Consideremos el siguiente modelo panel estático

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta w_{it} + c_i + u_{it}$$

donde $\mathbf{z}_{it}$ es estrictamente exógeno y w_{it} es secuencialmente exógena

$$E(u_{it}|\mathbf{z}_{it}, w_{it}, w_{it-1}, \ldots, w_{i1}, c_i) = 0$$

Sin embargo, w_{it} es influenciado por valor pasado de y_{it}

$$w_{it} = \mathbf{z}_{it}\xi + \rho_1 y_{it-1} + \psi c_i + r_{it}$$

Por ejemplo:
y_{it}: venta per capita de una vacuna en la ciudad i durante el año t
w_{it}: tasa de infección de un virus

- El modelo entonces puede ser usado para probar si la vacuna esta influenciada por la propagación del virus
- El efecto inobservado c_i contiene factores inobservables específicos a la ciudad
- La ecuación de w_{it} es una forma de capturar el hecho que la propagación del virus esta influenciada por la vacunación pasada

Marco general

Ejemplo: Modelo estático con *feedback*

Corroboremos exogeneidad estricta en el anterior modelo recursivo

$$E(w_{it+1}u_{it}) = \rho_1 E(y_{it}u_{it}) = E(u_{it}^2) > 0$$

Entonces, el supuesto de estricta exogeneidad se viola a menos que $\rho_1 = 0$

Algunas veces en aplicaciones de modelos de datos panel se incluyen variables rezagadas en lugar de contemporáneas, que pueden mitigar el problema pero no lo resuelve

La idea entonces es usar w_{it-1} en lugar de w_{it} en la ecuación principal de y_{it}, ya que pensamos que w_{it} y u_{it} están correlacionadas. El modelo será:

$$y_{it} = \mathbf{z}_{it}\boldsymbol{\gamma} + \delta w_{it-1} + c_i + u_{it}$$

En este caso $\mathbf{x}_{it+1} = (\mathbf{z}_{it+1}, w_{it})$, así que

$$\begin{aligned} E(\mathbf{x}'_{it+1}u_{it}) \quad & \Longrightarrow E(\mathbf{z}'_{it+1}u_{it}) = 0 \\ & \Longrightarrow E(w_{it}u_{it}) \neq 0 \end{aligned}$$

Así que la exogeneidad estricta falla y en estos caso la condición de exogeneidad secuencial es más razonable

Estimación bajo exogeneidad secuencial

La pregunta entonces que surge es: Qué estimador podemos aplicar bajo exogeneidad secuencial?

- Usar una transformación para eliminar c_i y buscar variables instrumentales
- El estimador de FE es inconsistente ya que el supuesto de exogeneidad estricta no se cumple
- Para modelos bajo el supuesto de exogeneidad secuencial, primeras diferencias es más atractivo

Estimación bajo exogeneidad secuencial

Sea el modelo

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

Se toman primeras diferencias para eliminar el efecto inobservable, esto es

$$(y_{it} - y_{it-1}) = (\mathbf{x}_{it} - \mathbf{x}_{it-1})\boldsymbol{\beta} + (u_{it} - u_{it-1})$$

$$\Delta y_{it} = \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}$$

Bajo el supuesto de exogeneidad secuencial se tiene que

$$E(\Delta \mathbf{x}_{it}' \Delta u_{it}) = -E(\mathbf{x}_{it}' u_{it}) \neq 0$$

Lo que muestra que el estimador de primeras diferencias es inconsistente, así que es necesario buscar potenciales instrumentos para $\Delta \mathbf{x}_{it}$

Estimación bajo exogeneidad secuencial

El supuesto de exogeneidad secuencial también implica que

$$E(\mathbf{x}_{is}' u_{it}) = 0$$

$$s = 1, \ldots, t; t = 1, \ldots, T$$

lo cual implica las siguientes condiciones de ortogonalidad

$$E(\mathbf{x}_{is}'(u_{it} - u_{it-1})) = E(\mathbf{x}_{is}' \Delta u_{it}) = 0$$

$$s = 1, \ldots, t-1; t = 2, \ldots, T$$

Por tanto, en el tiempo t, la disponibilidad de instrumentos en la ecuación de primeras diferencias están en el vector $\mathbf{x}_{it-1}^0$ donde

$$\mathbf{x}_{it}^0 = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{it})$$

El hecho que $\mathbf{x}_{it-1}^0$ esté incorrelacionado con Δu_{it} abre la posibilidad de varios procedimientos de estimación: 2SLS o GMM

Por ejemplo, un estimador que use $\Delta \mathbf{x}_{it-1}$ como instrumento para $\Delta \mathbf{x}_{it}$, se cumpliría que

$$E(\Delta \mathbf{x}_{it-1}' \Delta u_{it}) = 0$$

El modelo de primeras diferencias podría ser estimado por *pooled* 2SLS

Estimación bajo exogeneidad secuencial

Una característica general del procedimiento *pooled* 2SLS es que no usa todos los instrumentos disponibles en cada periodo de tiempo, por lo tanto, el procedimiento no es eficiente

El procedimiento óptimo es usar los instrumentos en una estimación por GMM

Entonces, la matriz de instrumentos es definida por

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{x}_{i1}^0 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}_{i2}^0 & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & & & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{x}_{iT-1}^0 \end{bmatrix} = \operatorname{diag}(\mathbf{x}_{i1}^0, \mathbf{x}_{i2}^0, \ldots, \mathbf{x}_{iT-1}^0)$$

Debido a la exogeneidad secuencial, el número de instrumentos validos incrementa con t

Como una cuestión práctica, la dimensión de columnas de $\mathbf{W}_i$ puede ser grande, especialmente cuando T es grande. Se sabe que el uso de muchas restricciones de sobre-identificación contribuye a que las propiedades de muestra finita de los GMM sean pobres, especialmente si muchos de los instrumentos son débiles

Lo recomendable, aunque no sea eficiente, es usar un par de rezagos como instrumentos

Modelos con variable dependiente rezagada

Un caso especial de modelos bajo restricciones de exogeneidad secuencial son los modelos autorregresivos. Aquí estudiamos el modelo AR(1) (Arellano y Bond, 1991), que tiene la siguiente estructura cuando no tiene variables explicativas adicionales

$$y_{it} = \rho y_{it-1} + c_i + u_{it}, t = 1, \ldots, T$$

y el supuesto de exogeneidad secuencial es

$$E(u_{it} | y_{it-1}, y_{it-2}, \ldots, y_{i0}, c_i) = 0$$

Una hipótesis interesante en este modelo es $H_0 : \rho_1 = 0$, que significa que, después de que la heterogeneidad inobservable ha sido controlado, y_{it-1} nos ayuda a predecir y_{it}.

Cuando $\rho_1 \neq 0$, decimos que y_{it} exhibe dependencia de estado (*state dependence*): el estado actual de la variable de resultado depende del estado del último período, incluso después de controlar por c_i

Modelos con variable dependiente rezagada

Estimación del modelo de datos panel dinámico

- En el tiempo t en la ecuación de primeras diferencias $\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}$, $t = 2, \ldots, T$, los instrumentos disponibles son $\mathbf{w}_{it} = (y_{i0}, \ldots, y_{t-2})$
- Anderson y Hsiao (1982): instrumento y_{it-2} o Δy_{it-2}
- Arellano y Bond (1991): GMM usando todas las variables instrumentales
- El estimador *Pooled* IV que usa y_{i0} como instrumento en $t = 2$ y luego ($y{*it-2},y*{it-3}$) para $t = 3, \ldots, T$. Este estimador es más eficiente que el enfoque Anderson y Hsiao, pero menos eficiente que el enfoque full GMM

Ejercicio aplicado en R: modelos dinámicos con datos de panel

En este ejercicio aplicado se van a utilizar los datos de Arellano y Bond (1991), en donde se estima una ecuación de empleo utilizando una muestra de empresas en UK. Se tiene información para 140 empresas en el periodo 1976 a 1984, en un panel desbalanceado.

Se considera la siguiente ecuación de empleo dinámica

$$
\begin{aligned}
ln(empleo_{it}) = \quad & \alpha_1 ln(empleo_{it-1}) + \alpha_2 ln(empleo_{it-2}) + \\
& \beta_1 ln(wage_{it}) + \beta_2 ln(wage_{it-1}) + \\
& \delta_1 ln(capital_{it}) + \delta_2 ln(capital_{it-1}) + \delta_3 ln(capital_{it-2}) + \\
& \gamma_1 ln(output_{it}) + \gamma_2 ln(output_{it-1}) + \gamma_3 ln(output_{it-2}) + e_{it}
\end{aligned}
$$

En el siguiente link se encuentra el código utilizado en R:

- Código en R