

Econometría de Datos de Panel

Maestrías en Economía y Econometría

Lecture 1

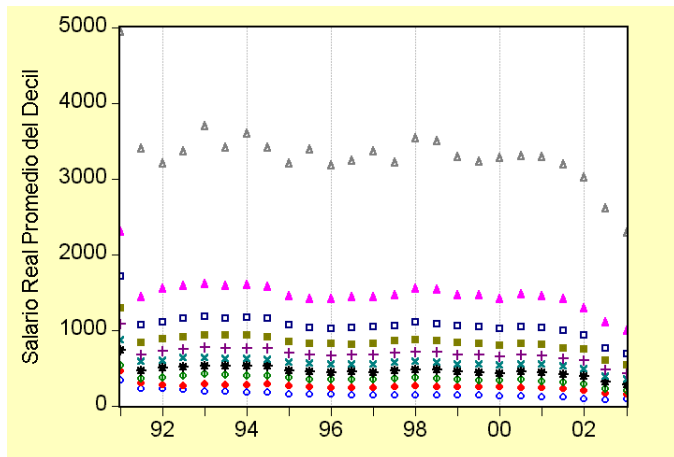
Agenda

- 1 Introducción
- 2 Teoría Asintótica
- 3 Modelos de Datos de Panel Lineales
- 4 Modelos de Efectos Fijos y Aleatorios
 - Modelo de Efectos Aleatorios
 - Modelo de Efectos Fijos
 - Transformación de Efectos Fijos ó Within Transformation
 - Modelo de Variables Binarias (LSDV)
 - Transformación de Diferencias Finitas
 - Transformación por Desviaciones Ortogonales
- 5 Two-Way Fixed Effects Model

Introducción

- Datos de Corte Transversal
 $j = 1, 2, \dots, N$
- Datos de Series Temporales
 $t = 1, 2, \dots, T$
- Datos de Panel
 $j = 1, 2, \dots, N$ y $t = 1, 2, \dots, T$

Introducción



- Datos Longitudinales. N es fijo y $T \rightarrow \infty$.
- Datos de Panel. $N \rightarrow \infty$ y T es fijo.

Convergencia en Probabilidad

- (1) Una secuencia de variables aleatorias $\{x_N : N = 1, 2, \dots\}$ converge en probabilidad a la constante a si para todo $\epsilon > 0$,

$$P[|x_N - a| > \epsilon] = 0 \text{ cuando } N \rightarrow \infty$$

En general, escribimos $x_N \xrightarrow{p} a$ y decimos que a es el plímite de x_N .

- (2) En el caso especial en que $a = 0$, también decimos que $\{x_N\}$ es $o_p(1)$ (o pequeña p uno). En este caso escribimos $x_N = o_p(1)$ ó $x_N \xrightarrow{p} 0$.

- (3) Una secuencia de variables aleatorias $\{x_N\}$ está limitada en probabilidad (bounded in probability) sí y solo sí para cada $\epsilon > 0$, existe un $b_\epsilon < \infty$ y un entero N_ϵ , tal que:

$$P[|x_N| \geq b_\epsilon] < \epsilon \text{ para todo } N > N_\epsilon$$

Convergencia en Probabilidad

- En este caso escribimos $x_N = O_p(1)$ ($\{x_N\}$ es O grande p uno).
- **Lema 1:** si $x_N \xrightarrow{p} a$, entonces $x_N = O_p(1)$
- (4) Una secuencia aleatoria $\{x_N : N = 1, 2, \dots\}$ es $o_p(N^\delta)$ para $\delta \in \Re$ si $N^{-\delta}x_N = o_p(1)$.
- **Lema 2:** si $w_N = o_p(1)$, $x_N = o_p(1)$, $y_N = O_p(1)$, y $z_N = O_p(1)$, entonces
 - (i) $w_N + x_N = o_p(1)$;
 - (ii) $y_N + z_N = O_p(1)$;
 - (iii) $y_N \times z_N = O_p(1)$;
 - (iv) $x_N \times z_N = o_p(1)$.
- Todas las definiciones anteriores se aplican elemento por elemento a secuencias de vectores y matrices.
- **Lema 3:** Sea $\{Z_N : N = 1, 2, \dots\}$ una secuencia de matrices $J \times K$ tal que $Z_N = o_p(1)$, y sea $\{x_N\}$ una secuencia de vectores aleatorios $J \times 1$ tal que $x_N = O_p(1)$. Entonces $Z_N'x_N = o_p(1)$.

Convergencia en Probabilidad

- **Lema 4 (Teorema de Slutsky):** Sea $g : \mathbb{R}^K \rightarrow \mathbb{R}^J$ una función continua en algún punto $c \in \mathbb{R}^K$. Sea $\{x_N : N = 1, 2, \dots\}$ una secuencia de vectores aleatorios $K \times 1$ tal que $x_N \xrightarrow{p} c$. Entonces $g(x_N) \xrightarrow{p} g(c)$ cuando $N \rightarrow \infty$. En otras palabras: $\text{plim } g(x_N) = g(\text{plim } x_N)$ si $g(\cdot)$ es continua en $\text{plim } x_N$.
- **Definición 1:** Sea $(\Omega, \mathfrak{F}, P)$ un espacio de probabilidad. Una secuencia de eventos $\{\Omega_N : N = 1, 2, \dots\} \subset \mathfrak{F}$ se dice que ocurre con probabilidad aproximándose a uno (w.p.a 1) sí y solo sí $P(\Omega_N) \rightarrow 1$ cuando $N \rightarrow \infty$.
- **Corolario 1:** Sea $\{Z_N : N = 1, 2, \dots\}$ una secuencia de matrices aleatorias $K \times K$, y sea A una matriz invertible no aleatoria $K \times K$. Si $Z_N \xrightarrow{p} A$ entonces:
 - (1) Z_N^{-1} existe w.p.a. 1
 - (2) $Z_N^{-1} \xrightarrow{p} A^{-1}$

Convergencia en Distribución

- **Definición 2:** Una secuencia de variables aleatorias $\{x_N : N = 1, 2, \dots\}$ converge en distribución a la variable aleatoria continua x sí y solo sí $F_N(\xi) \rightarrow F(\xi)$ cuando $N \rightarrow \infty$ para todo $\xi \in \mathfrak{R}$.
- Donde F_N es la función de distribución acumulada de x_N y F es la función de distribución acumulada de x . En este caso escribimos: $x_N \xrightarrow{d} x$
- **Definición 3:** Una secuencia de vectores aleatorios $\{x_N : N = 1, 2, \dots\}$ $K \times 1$ converge en distribución al vector aleatorio continuo x sí y solo sí para cualquier vector no aleatorio $K \times 1$, c tal que $c'c = 1$, $c'x_N \xrightarrow{d} c'x$ y escribimos $x_N \xrightarrow{d} x$.
- **Lema 5:** Si $x_N \xrightarrow{d} x$, donde x es cualquier vector aleatorio $K \times 1$, entonces $x_N = O_p(1)$.

Convergencia en Distribución

- **Lema 6 (Continuous mapping theorem):** Sea $\{x_N\}$ una secuencia de vectores aleatorios $K \times 1$, tal que $x_N \xrightarrow{d} x$. Si $g : \mathbb{R}^K \rightarrow \mathbb{R}^J$ es una función continua, entonces $g(x_N) \xrightarrow{d} g(x)$.
- **Corolario 2:** Si $\{z_N\}$ es una secuencia de vectores aleatorios $K \times 1$, tal que $z_N \xrightarrow{d} N(0, V)$. Entonces:
 - (1) Para cualquier matriz no aleatoria $K \times M$, $A: \quad A'z_N \xrightarrow{d} N(0, A'VA)$.
 - (2) $z_N'V^{-1}z_N \xrightarrow{d} \chi_K^2$.
- **Lema 7:** Sean $\{x_N\}$ y $\{z_N\}$ secuencias de vectores aleatorios $K \times 1$. Si $z_N \xrightarrow{d} z$ y $z_N - x_N \xrightarrow{p} 0$. Entonces $x_N \xrightarrow{d} z$.
- **Teorema 1:** Sea $\{w_j : j = 1, 2, \dots\}$ una secuencia de vectores aleatorios $G \times 1$, independientes, idénticamente distribuidos tal que $E(|w_{jg}|) < \infty$, $g = 1, 2, \dots, G$. Entonces la secuencia satisface la ley débil de los grandes números (WLLN): $N^{-1} \sum_{j=1}^N w_j \xrightarrow{p} \mu_w$ donde $\mu_w = E(w_j)$.

- **Teorema 2 (Lindeberg-Levy):** Sea $\{w_j : j = 1, 2, \dots\}$ una secuencia de vectores aleatorios $G \times 1$, independientes, idénticamente distribuidos tal que $E(|w_{jg}^2|) < \infty$, $g = 1, 2, \dots, G$ y $E(w_j) = 0$. Entonces la secuencia satisface el teorema central del límite (CLT): $N^{-1/2} \sum_{j=1}^N w_j \xrightarrow{d} N(0, B)$ donde $B = \text{Var}(w_j) = E(w_j w_j')$.

Modelos de Datos de Panel Lineales

- **Supuestos:** se dispone de una muestra aleatoria de la población. Esto es, tenemos observaciones de **CORTE TRANSVERSAL** $\{(X_j, y_j) : j = 1, 2, \dots, N\}$ que son independientes e idénticamente distribuídas.
- X_j es una matriz $T \times K$ que incluye como primera columna un vector $T \times 1$ de unos e y_j es un vector $T \times 1$.
- El modelo lineal multivariante para una muestra aleatoria de la población puede escribirse como:

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

$$y_j = X_j\beta + u_j, \quad j = 1, 2, \dots, N \quad (2)$$

donde β es el vector, $K \times 1$, de parámetros a estimar y u_j es un vector $T \times 1$ de errores no observables.

Modelos de Datos de Panel Lineales

- Dado el modelo planteado en (2) la pregunta relevante es **cuáles son los supuestos que se necesitan para poder estimar β consistentemente?**.
- Una posibilidad es asumir que x_{jt} y u_{jt} son ortogonales en sentido condicional:

$$E(u_{jt}|x_{jt}) = 0, \quad t = 1, 2, \dots, T \quad (3)$$

Esta relación recibe el nombre de **exogeneidad contemporánea de x_{jt}** .

- Es importante distinguir el supuesto de la ecuación (3) del más fuerte:

$$E(u_{jt}|x_{j1}, x_{j2}, \dots, x_{jT}) = 0, \quad t = 1, 2, \dots, T \quad (4)$$

denominado **exogeneidad estricta de las variables explicativas**, ya que implica que u_{jt} no está correlacionado con las variables explicativas en **NINGUNO** de los períodos temporales.

- La estimación consistente de β depende crucialmente de si se asume (3) ó (4).

Modelos de Datos de Panel Lineales

- En general, para poder estimar β en (2) en forma consistente por OLS necesitamos:
- **Supuesto 1:** $E(X_j' u_j) = 0$.
- El **supuesto 1** implica que $E(u_j) = 0$.
- En el caso de datos de panel $X_j' u_j = \sum_{t=1}^T x_{jt}' u_{jt}$, por lo tanto una condición natural para que el **supuesto 1** se satisfaga es $E(x_{jt}' u_{jt}) = 0$, $t = 1, 2, \dots, T$.
- Note que la ecuación anterior **NO** impone exogeneidad estricta. Bajo el **supuesto 1**, el vector β satisface:

$$\begin{aligned} E[X_j'(y_j - X_j\beta)] &= 0 \\ E(X_j'X_j)\beta &= E(X_j'y_j) \end{aligned} \tag{5}$$

Para poder estimar β necesitamos asumir que es el único vector $K \times 1$ que satisface (5).

Modelos de Datos de Panel Lineales

- **Supuesto 2.** $A \equiv E(X_j'X_j)$ es una matriz no aleatoria y no singular. Es decir, $\text{Rango}[E(X_j'X_j)] = K$.
- Bajo los **supuestos 1 y 2** podemos escribir:

$$\beta = [E(X_j'X_j)]^{-1}E(X_j'y_j) \quad (6)$$

- El principio de analogía sugiere estimar β con el análogo muestral de (6)

$$\hat{\beta} = \left(\frac{1}{N} \sum_{j=1}^N X_j'X_j \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N X_j'y_j \right) \quad (7)$$

- Note que en la ecuación anterior

$$\sum_{j=1}^N X_j'X_j = \sum_{j=1}^N \sum_{t=1}^T x_{jt}'x_{jt} \quad y \quad \sum_{j=1}^N X_j'y_j = \sum_{j=1}^N \sum_{t=1}^T x_{jt}'y_{jt}$$

Modelos de Datos de Panel Lineales

$$\begin{aligned}\hat{\beta} &= \left(\frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} x_{jt} \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} y_{jt} \right) \\ &= \beta + \left(\frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} x_{jt} \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} u_{jt} \right) \\ &\xrightarrow{p} \beta + \left(E \sum_{t=1}^T x'_{jt} x_{jt} \right)^{-1} \times \left(E \sum_{t=1}^T x'_{jt} u_{jt} \right) \\ &\xrightarrow{p} \beta + A^{-1} \text{por supuesto 2} \times 0 \text{ por supuesto 1}\end{aligned} \tag{8}$$

- La ecuación (8) recibe el nombre de **Pool OLS**.

Modelos de Datos de Panel Lineales

- Para realizar inferencia estadística lo único que necesitamos es la varianza asintótica del estimador.
- De la derivación anterior:

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} x_{jt} \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} u_{jt} \right) \quad (9)$$

- Como por el **Supuesto 1**, $E \left(\sum_{t=1}^T x'_{jt} u_{jt} \right) = 0$ lo único que necesitamos asumir para aplicar el CLT es: $E \left(\sum_{t=1}^T x'_{jt} u_{jt} u'_{jt} x_{jt} \right) < \infty$
- Con este supuesto adicional el CLT implica que:

$\frac{1}{\sqrt{N}} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} u_{jt} \xrightarrow{d} \text{Normal}(0, B)$, donde:

$$B \equiv E \left(\sum_{t=1}^T x'_{jt} u_{jt} u'_{jt} x_{jt} \right) = \text{Var} \left(\sum_{t=1}^T x'_{jt} u_{jt} \right)$$

Modelos de Datos de Panel Lineales

- Por lo tanto usando el **Supuesto 2** junto con el resultado anterior en (9), tenemos:

$$\begin{aligned}\sqrt{N}(\hat{\beta} - \beta) &\xrightarrow{d} A^{-1}\text{Normal}(0, B) \\ &\xrightarrow{d} \text{Normal}(0, A^{-1}BA^{-1})\end{aligned}\quad (10)$$

- Por lo tanto, la varianza asintótica del estimador de β es:

$$\text{AVar}(\hat{\beta}) = A^{-1}BA^{-1}/N \quad (11)$$

- Para realizar inferencia estadística necesitamos una estimación consistente de esta varianza asintótica:

$$\hat{A} = \frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} x_{jt}$$

Modelos de Datos de Panel Lineales

- Usando el principio de analogía, se puede obtener una estimación consistente de B como

$$\hat{B} = \frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} u_{jt} u'_{jt} x_{jt} \xrightarrow{p} B$$

- En la estimación anterior no conocemos los errores y por lo tanto deben ser reemplazados por los residuos del modelo. Esto es:

$$\hat{B} = \frac{1}{N} \sum_{j=1}^N \sum_{t=1}^T x'_{jt} \hat{u}_{jt} \hat{u}'_{jt} x_{jt} \xrightarrow{p} B$$

- Ahora, la última convergencia debe probarse. Note que:

$$\hat{B} = \frac{1}{N} \sum_{j=1}^N X'_j \hat{u}_j \hat{u}'_j X_j \quad y \quad \hat{u}_j = y_j - X_j \beta = u_j - X_j(\hat{\beta} - \beta)$$

Modelos de Datos de Panel Lineales

- Entonces,

$$\begin{aligned}\hat{B} &= \frac{1}{N} \sum_{j=1}^N X_j' [u_j - X_j(\hat{\beta} - \beta)] [u_j - X_j(\hat{\beta} - \beta)]' X_j \\ &= \frac{1}{N} \sum_{j=1}^N X_j' u_j u_j' X_j + \frac{1}{N} \sum_{j=1}^N X_j' X_j (\hat{\beta} - \beta) (\hat{\beta} - \beta)' X_j' X_j \\ &\quad + \frac{1}{N} \sum_{j=1}^N X_j' u_j (\hat{\beta} - \beta) X_j' X_j + \frac{1}{N} \sum_{j=1}^N X_j' X_j (\hat{\beta} - \beta) u_j' X_j \quad (12) \\ &\xrightarrow{p} B + o_p(1) + o_p(1) + o_p(1)\end{aligned}$$

Modelos de Datos de Panel Lineales

- Por lo tanto $AVar[\sqrt{N}(\hat{\beta} - \beta)]$ puede estimarse consistentemente con $\hat{A}^{-1}\hat{B}\hat{A}^{-1}$
- Y la varianza asintótica de $\hat{\beta}$ se estima con:

$$\hat{V} = \left(\sum_{j=1}^N X_j' X_j \right)^{-1} \left(\sum_{j=1}^N X_j' \hat{u}_j \hat{u}_j' X_j \right) \left(\sum_{j=1}^N X_j' X_j \right)^{-1} \quad (13)$$

- **Remark 1.** Bajo los **supuestos 1 y 2** se puede realizar inferencia estadística sobre β porque $\hat{\beta}$ se distribuye con una distribución normal con media β y matriz de varianzas y covarianzas dada por la ecuación (13).
- **Remark 2.** La raíz cuadrada de los elementos de la diagonal principal de (13) se reportan usualmente como los errores estándar asintóticos de los estimadores.

Modelos de Datos de Panel Lineales

- **Remark 3.** El cociente $t, \hat{\beta}_j/\text{se}(\hat{\beta}_j)$, tiene distribución normal bajo la hipótesis nula $H_0 : \beta_j = 0$. Usualmente estos estadísticos son tratados siguiendo una distribución t -Student con $N \times T - K$ grados de libertad, que es una aproximación asintóticamente válida.
- **Remark 4.** La ecuación (13) es válida sin haber hecho ningún supuesto sobre el momento de orden dos de los errores.
- El **Remark 4** implica que la matriz $T \times T$ de varianzas y covarianzas de los errores no está restringida de ninguna manera (i.e. $\Omega = E(u_j u_j')$).
- Ω no restringida permite cualquier tipo de correlación serial y varianzas de los disturbios que varíen en el tiempo.
- Inferencia sobre múltiples hipótesis puede realizarse utilizando el estadístico de Wald.

Modelos de Datos de Panel Lineales

- El estadístico (robusto) de Wald para contrastar la hipótesis nula $H_0 : R\beta = r$, donde R es $Q \times K$ con rango Q y r es $Q \times 1$ sigue la expresión usual:

$$W = (R\hat{\beta} - r)'[R\hat{V}R']^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi_Q^2 \quad (14)$$

- El test de Wald permite contrastar cualquier hipótesis sobre β sin tener que asumir homocedasticidad ni independencia serial de los errores.
- Para aplicar los estadísticos usuales de OLS para la estimación pool, necesitamos asumir homocedasticidad y ausencia de correlación serial en el tiempo. Las formas más débiles de estos supuestos son:

Supuesto 3: (a) $E(u_{jt}^2 x'_{jt} x_{jt}) = \sigma^2 E(x'_{jt} x_{jt})$, $t = 1, 2, \dots, T$ y $\sigma^2 = E(u_{jt}^2)$ para todo t . (b) $E(u_{jt} u_{js} x'_{jt} x_{js}) = 0$, $t \neq s$, $t, s = 1, 2, \dots, T$.

Modelos de Datos de Panel Lineales

- El supuesto 3 implica que,

$$B \equiv E \left(\sum_{t=1}^T x'_{jt} u_{jt} u'_{jt} x_{jt} \right) = \sigma^2 E \left(\sum_{t=1}^T x'_{jt} x_{jt} \right) = \sigma^2 A$$

y por lo tanto, $\text{AVar}(\hat{\beta}) = \sigma^2 A/N$

- El estimador apropiado es entonces,

$$\hat{V} = \widehat{\text{AVar}}(\hat{\beta}) = \hat{\sigma}^2 \left(\sum_{j=1}^N \sum_{t=1}^T x'_{jt} x_{jt} \right)^{-1} \quad (15)$$

- En la ecuación anterior, $\hat{\sigma}^2$ es el estimador usual de la varianza de los errores del modelo estimado por Pool OLS.
- Si el supuesto 3 se satisface, los estadísticos t y F son válidos.
- Si el supuesto 3 no se satisface, una alternativa a OLS es mínimos cuadrados generalizados (GLS).

Modelos de Datos de Panel Lineales

- Para poder establecer la consistencia de GLS necesitamos reforzar el **supuesto 1**.
- **Supuesto 1'**. $E(X_j \otimes u_j) = 0$. Esto es, cada elemento de u_j no está correlacionado con cada elemento de X_j .
- Una condición suficiente para que se cumpla el **supuesto 1'** es $E(u_j|X_j) = 0$.
- En lugar del **supuesto 2**, necesitamos
- **Supuesto 2'**. Ω es positiva definida y $E(X_j' \Omega^{-1} X_j)$ no es singular.
- Pre-multiplicando el modelo (2) por $\Omega^{-1/2}$ tenemos:

$$\begin{aligned}\Omega^{-1/2} y_j &= \Omega^{-1/2} X_j \beta + \Omega^{-1/2} u_j \\ y_j^* &= X_j^* \beta + u_j^*\end{aligned}\tag{16}$$

- Ahora $E(u_j^* u_j^{*'}) = I_T$.

Modelos de Datos de Panel Lineales

- Aplicando OLS a (16) obtenemos:

$$\begin{aligned}\tilde{\beta} &= \left(\frac{1}{N} \sum_{j=1}^N X_j^{*'} X_j^* \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N X_j^{*'} y_j^* \right) \\ &= \left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} X_j \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} y_j \right)\end{aligned}\quad (17)$$

$$= \beta + \left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} X_j \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} u_j \right)\quad (18)$$

- Por la ley de los grandes números (WLLN)

$$\left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} X_j \right) \xrightarrow{P} E(X_j' \Omega^{-1} X_j) \equiv A$$

Modelos de Datos de Panel Lineales

- Ahora, por el **Supuesto 2'** y el **Lema 4 (teorema de Slutsky)** tenemos que:

$$\left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} X_j \right)^{-1} \xrightarrow{p} A^{-1}$$

- Por la ley de los grandes números (WLLN)

$$\left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} u_j \right) \xrightarrow{p} E (X_j' \Omega^{-1} u_j)$$

- Bajo el **Supuesto 1'**: $E (X_j' \Omega^{-1} u_j) = 0$

- Proof:**

$$\begin{aligned} \text{Vec} [E (X_j' \Omega^{-1} u_j)] &= E [u_j' \otimes X_j'] \text{Vec} (\Omega^{-1}) \\ &= E [(u_j \otimes X_j)'] \text{Vec} (\Omega^{-1}) = 0 \end{aligned}$$

Modelos de Datos de Panel Lineales

- Usando la ecuación (18) tenemos,

$$\sqrt{N}(\tilde{\beta} - \beta) = \left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} X_j \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \Omega^{-1} u_j \right) \quad (19)$$

- Como vimos, por la WLLN y el **Supuesto 1'**, $E(X_j' \Omega^{-1} u_j) = 0$ por lo tanto, por el CLT

$$\left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \Omega^{-1} u_j \right) \xrightarrow{d} \text{Normal}(0, B)$$

where $B = E(X_j' \Omega^{-1} u_j u_j' \Omega^{-1} X_j)$.

Modelos de Datos de Panel Lineales

- Usando los resultados anteriores, tenemos

$$\begin{aligned}\sqrt{N}(\tilde{\beta} - \beta) &= \left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} X_j \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \Omega^{-1} u_j \right) \\ &\xrightarrow{d} A^{-1} \text{Normal}(0, B) \\ &\xrightarrow{d} \text{Normal}(0, A^{-1} B A^{-1})\end{aligned}$$

- Por lo tanto $A\text{Var}(\tilde{\beta}) = A^{-1} B A^{-1} / N$.
- Para poder estimar el modelo por GLS necesitamos conocer Ω . En la práctica rara vez conocemos esta matriz y el método no puede ser utilizado empíricamente.
- Sin embargo, existe un método que es asintóticamente equivalente conocido como **mínimos cuadrados generalizados estimados** o **FGLS**.

Modelos de Datos de Panel Lineales

- En FGLS reemplazamos Ω con una estimación consistente.
- En general se utiliza el siguiente procedimiento.
 - (1) Obtenga el estimador Pool OLS de β , $\hat{\beta}$, from (8).
 - (2) Obtenga los residuos $\hat{u}_j = y_j - X_j \hat{\beta}$.
 - (3) Estime Ω con $\tilde{\Omega} = \sum_{j=1}^N \hat{u}_j \hat{u}_j'$.
- Note que, $\hat{u}_j = u_j - X_j(\hat{\beta} - \beta)$, por lo tanto

$$\begin{aligned}\hat{u}_j \hat{u}_j' &= u_j u_j' - u_j (\hat{\beta} - \beta)' X_j' - X_j (\hat{\beta} - \beta) u_j' \\ &\quad + X_j (\hat{\beta} - \beta) (\hat{\beta} - \beta)' X_j'\end{aligned}$$

Modelos de Datos de Panel Lineales

- El segundo término del lado derecho puede escribirse como

$$\begin{aligned}\frac{1}{N} \sum_{j=1}^N \text{Vec}[u_j(\hat{\beta} - \beta)' X_j'] &= \frac{1}{N} \sum_{j=1}^N (X_j \otimes u_j) \text{Vec}(\hat{\beta} - \beta)' \\ &\xrightarrow{p} E(X_j \otimes u_j) \times o_p(1)\end{aligned}$$

- El último término del lado derecho puede escribirse como:

$$\begin{aligned}\frac{1}{N} \sum_{j=1}^N X_j(\hat{\beta} - \beta)(\hat{\beta} - \beta)' X_j' &= \frac{1}{N} \sum_{j=1}^N (X_j \otimes X_j') \text{Vec}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' \\ &\xrightarrow{p} O_p(1) \times o_p(1)\end{aligned}$$

Modelos de Datos de Panel Lineales

- Por lo tanto

$$\tilde{\Omega} = \sum_{j=1}^N u_j u_j' + o_p(1) \implies \tilde{\Omega} \xrightarrow{p} \Omega \quad (20)$$

- El estimador FGLS de β es entonces:

$$\begin{aligned} \hat{\beta} &= \left(\frac{1}{N} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} X_j \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} y_j \right) \\ &= \beta + \left(\frac{1}{N} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} X_j \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} u_j \right) \end{aligned} \quad (21)$$

Modelos de Datos de Panel Lineales

- De la última expresión tenemos

$$\sqrt{N}(\hat{\beta} - \beta) = \left(\frac{1}{N} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} X_j \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} u_j \right)$$

- Ahora,

$$\begin{aligned} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} u_j \right) - \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \Omega^{-1} u_j \right) &= \\ \left[\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' (\tilde{\Omega}^{-1} - \Omega^{-1}) u_j \right] &= \\ \left[\frac{1}{\sqrt{N}} \sum_{j=1}^N (u_j \otimes X_j)' \right] \text{Vec}(\tilde{\Omega}^{-1} - \Omega^{-1}) &\xrightarrow{p} O_p(1) \times o_p(1) \end{aligned}$$

Modelos de Datos de Panel Lineales

- Por lo tanto

$$\left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} u_j \right) = \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \Omega^{-1} u_j \right) + o_p(1)$$

- Usando el mismo argumento

$$\left(\frac{1}{N} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} X_j \right) = \left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} X_j \right) + o_p(1)$$

- Utilizando los dos resultados anteriores tenemos que

$$\begin{aligned} \sqrt{N}(\hat{\beta} - \beta) &= \left(\frac{1}{N} \sum_{j=1}^N X_j' \Omega^{-1} X_j \right)^{-1} \left(\frac{1}{\sqrt{N}} \sum_{j=1}^N X_j' \Omega^{-1} u_j \right) + o_p(1) \\ &= \sqrt{N}(\tilde{\beta} - \beta) + o_p(1) \implies \sqrt{N}(\hat{\beta} - \tilde{\beta}) = o_p(1) \end{aligned}$$

y ambos estimadores, SGLS y FGLS, son asintóticamente equivalentes (\sqrt{N} equivalentes).

Modelos de Datos de Panel Lineales

- **Remark 5:** Empíricamente, la equivalencia asintótica de los estimadores de GLS y FGLS implica que para realizar inferencia estadística sobre β usando FGLS, no hay que preocuparse de que $\tilde{\Omega}$ sea un estimador de Ω .
- **Resumen:** Bajo los supuestos 1' y 2',

$$\sqrt{N}(\hat{\beta} - \beta) \xrightarrow{d} \text{Normal}(0, A^{-1}BA^{-1})$$

- Bajo FGLS un estimador consistente de A es:

$$\tilde{A} = \left(\frac{1}{N} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} X_j \right)$$

- Un estimador consistente de B es,

$$\tilde{B} = \frac{1}{N} \sum_{j=1}^N X_j' \tilde{\Omega}^{-1} \tilde{u}_j \tilde{u}_j' \tilde{\Omega}^{-1} X_j \quad \text{con} \quad \tilde{u}_j = y_j - X_j \hat{\beta}.$$

Modelos de Datos de Panel Lineales

- Usando los resultados anteriores una estimación de la varianza asintótica del estimador FGLS es:

$$\hat{V} = \widehat{AVar}(\hat{\beta}) = \tilde{A}^{-1} \tilde{B} \tilde{A}^{-1} / N = \left(\sum_{j=1}^N X_j' \tilde{\Omega}^{-1} X_j \right)^{-1} \left(\sum_{j=1}^N X_j' \tilde{\Omega}^{-1} \tilde{u}_j \tilde{u}_j' \tilde{\Omega}^{-1} X_j \right) \left(\sum_{j=1}^N X_j' \tilde{\Omega}^{-1} X_j \right)^{-1} \quad (22)$$

- La especificación anterior es la más general que se pueda tener en términos de los supuestos acerca de la estimación de la varianza asintótica.
- Supuesto 3'.** $E(X_j' \Omega^{-1} u_j u_j' \Omega^{-1} X_j) = E(X_j' \Omega^{-1} X_j)$, con $\Omega = E(u_j u_j')$.
- Bajo los **supuestos 1', 2' y 3'**, la varianza asintótica del estimador FGLS es:

$$AVar(\hat{\beta}) = A^{-1} / N = [E(X_j' \Omega^{-1} X_j)]^{-1} / N$$

Modelos de Datos de Panel Lineales

- Uno obtiene un estimador consistente de la varianza asintótica usando un estimador consistente de A .

$$\widehat{AVar}(\hat{\beta}) = \hat{A}^{-1}/N = \left(\sum_{j=1}^N x_j' \tilde{\Omega}^{-1} x_j \right)^{-1} \quad (23)$$

- Los errores estándar asintóticos de los coeficientes estimados se obtienen en forma usual utilizando la raíz cuadrada de la diagonal principal de (22) o de (23).
- Contrastes de múltiples hipótesis pueden realizarse utilizando el estadístico de Wald definido anteriormente en la ecuación (14).
- La única decisión importante es la elección de la estimación de la varianza asintótica correcta.
- El estadístico de Wald estándar usa (23) mientras que el robusto utiliza (14).

Contraste de Correlación Serial en POLS

- Supongamos correlación serial de primer orden: $u_{jt} = \alpha_1 u_{jt-1} + e_{jt}$, con $E(e_{jt} | X_{jt}, u_{jt-1}, \dots) = 0$.
- LM Test para $H_0 : \alpha_1 = 0$
 - (1) Estime por POLS $y_{jt} = x_{jt}\beta + u_{jt}$ y obtenga \hat{u}_{jt}
 - (2) Estime por POLS $\hat{u}_{jt} = x_{jt}\beta + \alpha_1 \hat{u}_{jt-1} + e_{jt}$ (ax)
 - (3) $LM = \text{Obs} \times R_{ax}^2 \sim \chi_1^2$.
 - (4) Regla de decisión: si el valor-p(LM) < nivel de error del test, entonces rechazar H_0 .
- El test tiene las generalizaciones usuales para correlación de mayor orden.

Contraste de Heterocedasticidad en POLS

- Este test es válido si se asume $E(u_{jt}|x_{jt}) = 0$, $t = 1, 2, \dots, T$.
- La hipótesis nula es entonces $E(u_{jt}^2|x_{jt}) = \sigma^2$, $t = 1, 2, \dots, T$.
- Bajo H_0 , u_{jt}^2 no está correlacionado con ninguna función de x_{jt} . Denotemos por h_{jt} a un vector de dimensión $1 \times Q$ de funciones no constantes de x_{jt} .
 - (1) Estime por POLS $y_{jt} = x_{jt}\beta + u_{jt}$ y obtenga \hat{u}_{jt} y \hat{u}_{jt}^2 .
 - (2) Estime por POLS \hat{u}_{jt}^2 sobre una constante y h_{jt} y obtenga el R_{ax}^2 .
 - (3) $LM = N \times T \times R_{ax}^2 \sim \chi_Q^2$.
 - (4) Regla de decisión: si el valor-p(LM) < nivel de error del test, entonces rechazar H_0 .

Modelos de Datos de Panel Lineales

- Hasta ahora, hemos supuesto, como mínimo que no existía correlación entre el error del período y las variables explicativas del modelo.
- Para ciertos datos de panel este supuesto es demasiado fuerte. Hay varios casos en los que uno debiera esperar una correlación entre variables observables y no observables.
- Un ejemplo clásico de este problema es el del **error de medición**. Si la variable explicativa que observamos no se mide correctamente el error de la ecuación contendrá este error de medición y por lo tanto estará correlacionado con la variable explicativa mal medida.
- Otro ejemplo posible es el de **variable omitida**.
- Justamente, uno de los usos más frecuentes de los datos de panel, es el de resolver los problemas de variables omitidas.
- Es fácil ver como los datos de panel nos pueden ayudar a resolver el problema de variables omitidas.

Modelos de Datos de Panel Lineales

- Sean $x = (x_1, x_2, \dots, x_k)$ e y variables aleatorias observables; y c una variable aleatoria no observable.
- Usualmente, estamos interesados en estimar los efectos parciales de las variables explicativas observables sobre la variable dependiente.
- **El Problema:** asumiendo un modelo lineal,

$$E(y|x, c) = \beta_0 + x\beta + c$$

- Estamos interesados en el vector β .
- Si $\text{Cov}(x_j, c) \neq 0$ para algún j , no podemos estimar consistentemente el vector β ni con OLS ni con GLS.
- En el contexto de datos de panel c recibe el nombre de **componente no observable, efecto no observable o heterogeneidad no observable**.
- La solución al problema de variables omitidas en panel consiste simplemente en transformar el modelo para eliminar c y luego estimar.

Modelos de Efectos Fijos y Aleatorios

- Consideremos un modelo para variables observadas a través de unidades de corte transversal durante varios períodos de tiempo. β es el parámetro que estamos interesados en estimar y c_i es un efecto no observado, invariante en el tiempo, denominado **efecto individual**, **heterogeneidad individual** ó **heterogeneidad no observada**:

$$y_{it} = c_i + x_{it}\beta + u_{it} \quad (24)$$

donde x_{it} es $1 \times K$ y u_{it} es el error idiosincrático. Este modelo se denomina **modelo de efectos no observables**.

- Tradicionalmente, existen dos modelos basados en la discusión acerca de si c_i puede tratarse como un **efecto aleatorio** o como un **efecto fijo**.
- Estas discusiones se centraban en si el efecto individual era una variable aleatoria o podía considerarse como un parámetro a ser estimado.

Modelos de Efectos Fijos y Aleatorios

- En el análisis de panel tradicional c_i se llama un efecto aleatorio cuando se lo trata como variable aleatoria y un efecto fijo cuando se lo trata como parámetro a ser estimado.
- Modernamente, la discusión ha cambiado y lo que se discute es básicamente si el efecto no observable está o no correlacionado con las variables explicativas observables.
- Ahora, efecto aleatorio es sinónimo de ausencia de correlación entre las variables explicativas observables y el efecto no observable:
$$\text{Cov}(x_{it}, c_i) = 0, t = 1, 2, \dots, T.$$
- En los trabajos empíricos cuando se dice que el modelo tiene un efecto aleatorio individual es porque se está asumiendo que no existe correlación entre las variables explicativas observables y el efecto no observable
- Similarmente, el término efecto fijo, no quiere decir que c_i se trate como no aleatorio, sino que implica que se permite la correlación entre c_i y x_{it} .

Exogeneidad Estricta

- $E(y_{it}|x_{i1}, x_{i2}, \dots, x_{iT}, c_i) = E(y_{it}|x_{it}, c_i) = x_{it}\beta + c_i$
con $t = 1, 2, \dots, T$.
- Cuando la ecuación anterior se satisface, se dice que las **variables explicativas son estrictamente exógenas condicionando en el efecto no observable**.
- La condición de exogeneidad estricta puede establecerse en términos de los errores idiosincráticos usando el modelo (24),

$$E(u_{it}|x_{i1}, x_{i2}, \dots, x_{iT}, c_i) = 0, \quad t = 1, 2, \dots, T. \quad (25)$$

- La ecuación (25) implica que las variables explicativas en cada período de tiempo no están correlacionadas con el error idiosincrático en cada período de tiempo:

$$E(x'_{is} u_{it}) = 0, \quad s, t = 1, 2, \dots, T \quad (26)$$

Exogeneidad Estricta

- Este supuesto de exogeneidad es mucho más fuerte que el de ausencia de correlación contemporánea
- $E(x'_{it}u_{it}) = 0, \quad t = 1, 2, \dots, T.$
- No obstante, note que (25) permite correlación arbitraria entre las variables explicativas observables y el efecto individual no observable.
- Estimación del modelo de efectos no observables con POLS
- Re-escribamos el modelo (24) como:

$$y_{it} = x_{it}\beta + v_{it} \quad t = 1, 2, \dots, T \quad (27)$$

donde $v_{it} = c_i + u_{it}$, $t = 1, 2, \dots, T$ son los **errores compuestos**.

- De acuerdo a lo que hemos visto, sabemos que la ecuación anterior puede estimarse por OLS y obtener estimadores consistentes si $E(x'_{it}v_{it}) = 0, \quad t = 1, 2, \dots, T.$

Modelos de Panel Lineales

- La última condición implica que estamos asumiendo que:
- $E(x'_{it}u_{it}) = 0$ y $E(x'_{it}c_i) = 0$, $t = 1, 2, \dots, T$.
- Note que el supuesto restrictivo aquí es la segunda condición.
- En modelos de panel dinámicos la segunda condición no puede cumplirse porque la variable dependiente rezagada (i.e. y_{it-1}) y c_i están necesariamente correlacionadas.
- Note que aún cuando las condiciones anteriores se satisfagan los errores compuestos estarán correlacionados debido a la presencia de c_i en cada período temporal.
- Una consecuencia del punto anterior es que para realizar inferencia usando POLS se deben calcular los errores estándar de los coeficientes usando la ecuación (13).
- Otra consecuencia que afectará nuestro análisis más adelante es que como v_{it} depende de c_i para todo t , la correlación entre v_{it} y v_{it-s} , $s > 0$ no decrece con s .
- En el lenguaje de las series temporales: v_{it} no tiene dependencia débil a través del tiempo.

Agenda

- 1 Introducción
- 2 Teoría Asintótica
- 3 Modelos de Datos de Panel Lineales
- 4 Modelos de Efectos Fijos y Aleatorios**
 - Modelo de Efectos Aleatorios
 - Modelo de Efectos Fijos
 - Transformación de Efectos Fijos ó Within Transformation
 - Modelo de Variables Binarias (LSDV)
 - Transformación de Diferencias Finitas
 - Transformación por Desviaciones Ortogonales
- 5 Two-Way Fixed Effects Model

Modelo de Efectos Aleatorios

- Como en el caso de POLS, los métodos de efectos aleatorios ponen a c_i en el término de error.
- En general el análisis de efectos aleatorios necesita supuestos más fuertes que POLS: exogeneidad estricta más ortogonalidad entre c_i y x_{it} .
- **Supuesto RE.1:**
 - (a) $E(u_{it}|X_i, c_i) = 0$ y
 - (b) $E(c_i|X_i) = 0$ con $X_i = (x_{i1}, x_{i2}, \dots, x_{iT})$
- Necesitamos (a) porque RE estima por GLS, debido a la correlación serial y como vimos antes, GLS necesita exogeneidad estricta para conseguir estimadores consistentes.
- Bajo RE.1 podemos escribir el modelo como:

$$y_{it} = x_{it}\beta + v_{it} \quad (28)$$

Modelo de Efectos Aleatorios

- En (28) $E(v_{it}|X_i) = 0$, $t = 1, 2, \dots, T$.
- Note que esta última ecuación implica que $\{x_{it}, t = 1, 2, \dots, T\}$ satisface el supuesto de exogeneidad estricta 1' en POLS.
- Por lo tanto podemos aplicar GLS tomando en cuenta la estructura de la matriz de varianzas y covarianzas del error.
- Escribamos el modelo *stacking* sobre T .

$$y_i = X_i\beta + v_i, \quad v_i = c_i J_T + u_i$$

donde J_T es un vector $T \times 1$ de unos.

- Definamos la matriz de varianzas y covarianzas de los errores del modelo como: $\Omega = E(v_i v_i')$, una matriz $T \times T$ positiva definida.
- Recuerde que esta matriz **es la misma para todo i** por el supuesto de muestra aleatoria.

Modelo de Efectos Aleatorios

- Para obtener estimadores consistentes por GLS necesitamos: **Supuesto RE.2:**
 $\text{rango} E(X_i' \Omega^{-1} X_i) = K.$
- Aplicando los resultados vistos antes para GLS sabemos que podemos obtener estimadores consistentes y asintóticamente normales con $N \longrightarrow \infty$ y usando una matriz Ω general.
- Pero podemos hacerlo mejor porque conocemos la estructura de los errores.
- Los supuestos tradicionales de RE son:
 - (i) $E(u_{it}^2) = \sigma_u^2, \quad t = 1, 2, \dots, T.$
 - (ii) $E(u_{it} u_{is}) = 0, \quad \forall t \neq s.$

En este caso:

$$\begin{aligned}\Omega = E(v_i v_i') &= E[(c_i J_T + u_i)(c_i J_T + u_i)'] \\ &= E\{c_i^2 J_T J_T' + c_i u_i J_T' + c_i J_T u_i' + u_i u_i'\} \\ &= \sigma_c^2 J_T J_T' + \sigma_u^2 I_T\end{aligned}$$

Modelo de Efectos Aleatorios

- De la ecuación anterior tenemos:

$$\Omega = E(v_i v_i') = \begin{bmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \cdots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \cdots & \sigma_c^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_c^2 & \sigma_c^2 & \cdots & \sigma_c^2 + \sigma_u^2 \end{bmatrix} \quad (29)$$

- Cuando Ω tiene la forma (29) se dice que tiene la estructura de efectos aleatorios.
- Note que en este caso Ω depende solo de dos parámetros σ_c^2 y σ_u^2 , independientemente del tamaño de T .
- Para obtener estimadores eficientes necesitamos que: (iii) $E(v_i v_i' | X_i) = E(v_i v_i')$
- El siguiente supuesto implica (i), (ii) e (iii)
- Supuesto RE.3:**
 - (a) $E(u_i u_i' | X_i, c_i) = \sigma_u^2 I_T$
 - (b) $E(c_i^2 | X_i) = \sigma_c^2$.

Bajo RE.3, Ω tiene la forma (29).

Modelo de Efectos Aleatorios

- **Implementación:** Necesitamos un estimador consistente de Ω .
- Asumamos que tenemos estimadores consistentes de σ_c^2 y σ_u^2 entonces:

$$\hat{\Omega} = \hat{\sigma}_c^2 J_T J_T' + \hat{\sigma}_u^2 I_T \quad (30)$$

- El estimador FGLS que usa (30) se conoce como **estimador de efectos aleatorios**.

●

$$\hat{\beta}^{RE} = \left(\sum_{i=1}^N x_i' \hat{\Omega}^{-1} x_i \right)^{-1} \left(\sum_{i=1}^N x_i' \hat{\Omega}^{-1} y_i \right)$$

- El estimador anterior es consistente bajo **RE.1 y RE.2**.

$$\hat{\beta}^{RE} = \beta + \left(\sum_{i=1}^N x_i' \hat{\Omega}^{-1} x_i \right)^{-1} \left(\sum_{i=1}^N x_i' \hat{\Omega}^{-1} v_i \right)$$

Modelo de Efectos Aleatorios

- Bajo el supuesto RE.3, el estimador de efectos aleatorios es eficiente.
- La matriz usual de varianzas-covarianzas de FGLS (23) es válida pero con $\hat{\Omega}$ dada por (30) en lugar de $\tilde{\Omega}$.

$$\widehat{\text{AVar}}(\hat{\beta}) = \left(\sum_{j=1}^N X_j' \hat{\Omega}^{-1} X_j \right)^{-1} \quad (31)$$

- Para poder implementar FGLS necesitamos las estimaciones consistentes de σ_c^2 y σ_u^2 . Para esto definamos $\sigma_v^2 = \sigma_c^2 + \sigma_u^2$.
- Bajo el supuesto RE.3(a), $\sigma_v^2 = T^{-1} \sum_{t=1}^T E(v_{it}^2)$ para todo i .
- Por lo tanto, un estimador consistente de σ_v^2 es:

$$\hat{\sigma}_v^2 = \frac{1}{NT - K} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2$$

Donde \hat{v}_{it} son los residuos de POLS.

Modelo de Efectos Aleatorios

- Para encontrar un estimador consistente de σ_c^2 , recuerde que $\sigma_c^2 = E(v_{it}v_{is})$, para todo $t \neq s$.
- Un estimador consistente es entonces:

$$\hat{\sigma}_c^2 = \frac{1}{[NT(T-1) - 1]/2 - K} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{v}_{it} \hat{v}_{is} \quad (32)$$

- Con estos resultados podemos estimar consistentemente: $\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2$.
- En la práctica, la ecuación para estimar σ_c^2 no garantiza una estimación POSITIVA.
- Si la estimación da negativa entonces eso es un signo de que existe correlación negativa en u_{it} lo que implica que RE.3 no se cumple.
- En este caso debieramos estimar FGLS sin restricciones.

Modelo de Efectos Aleatorios

- Si el supuesto RE.3 no se cumple es importante poder realizar inferencia estadística sin ese supuesto.
- Para ello, simplemente utilizamos la estimación robusta de la matriz de varianzas y covarianzas dada por la ecuación (22), reemplazando \tilde{u}_i por $\hat{v}_i = y_i - X_i \hat{\beta}^{RE}$, $i = 1, 2, \dots, N$

$$\hat{V} = \left(\sum_{i=1}^N X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \left(\sum_{i=1}^N X_i' \hat{\Omega}^{-1} \hat{v}_i \hat{v}_i' \hat{\Omega}^{-1} X_i \right) \left(\sum_{i=1}^N X_i' \hat{\Omega}^{-1} X_i \right)^{-1} \quad (33)$$

- Los errores estándar robustos se obtienen de la raíz cuadrada de la diagonal principal de (33), y el test de Wald robusto se obtiene con la fórmula:

$$W = (R \hat{\beta}^{RE} - r)' [R \hat{V} R']^{-1} (R \hat{\beta}^{RE} - r) \xrightarrow{d} \chi_Q^2 \quad (34)$$

donde \hat{V} es la matriz de varianzas y covarianzas estimada en forma robusta.

Modelo de Efectos Aleatorios

- Si los errores idiosincráticos $\{u_{it} : t = 1, 2, \dots, T\}$ son heterocedásticos y/o tienen correlación serial, se debe utilizar un estimador de Ω más general:

$$\hat{\Omega} = N^{-1} \sum_{i=1}^N \hat{\mathbf{v}}_i \hat{\mathbf{v}}_i'$$

donde $\hat{\mathbf{v}}_i$ son los residuos de POLS.

- Con N grande el estimador de FGLS más general es tan eficiente como RE.
- El estimador de FGLS más general es más eficiente que RE si $E(\mathbf{v}_i \mathbf{v}_i' | \mathbf{X}_i) = \Omega$, pero Ω no tiene la estructura de efectos aleatorios.
- Por qué entonces no se usa siempre el modelo más general?
- Históricamente, la estructura de la matriz de varianzas y covarianzas de RE se consideró sinónimo de efectos no observables.

Modelo de Efectos Aleatorios

- Contraste por la presencia de un efecto no observable.
- Si los **supuestos RE.1-RE.3** se cumplen pero **no existe un efecto no observable**, entonces POLS es eficiente y todos los estadísticos asociados a POLS son asintóticamente válidos.
- La ausencia de un efecto no observable es estadísticamente equivalente a $H_0 : \sigma_c^2 = 0$.
- El estadístico de contraste se basa en (32) y en la distribución asintótica de

$$N^{-1/2} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{v}_{it} \hat{v}_{is} \quad (35)$$

que es esencialmente el estimador de σ_c^2 escalado por $N^{-1/2}$.

- Por el supuesto de exogeneidad estricta la distribución de (35) es la misma con los residuos de POLS que con los errores verdaderos.

Modelo de Efectos Aleatorios

- $N^{-1/2} \sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T v_{it} v_{is}$ tiene distribución normal con varianza dada por $E \left(\sum_{t=1}^{T-1} \sum_{s=t+1}^T v_{it} v_{is} \right)$
- Haciendo el cociente entre (35) y su error estándar tenemos un estadístico de contraste con distribución normal estandar.

$$Z = \frac{\sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{v}_{it} \hat{v}_{is}}{\left(\sum_{i=1}^N \left(\sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{v}_{it} \hat{v}_{is} \right)^2 \right)^{1/2}} \quad (36)$$

- El estadístico (36) tiene la capacidad de detectar muchas formas de correlación serial en el error compuesto.
- Tradicionalmente, el estadístico de contraste utilizado para detectar la presencia de efectos no observables es el estadístico del multiplicador de Lagrange (Breusch-Pagan, 1980).

Modelo de Efectos Aleatorios

- La hipótesis nula del test de Breusch-Pagan es la misma: $H_0 : \sigma_c^2 = 0$, y el estadístico de contraste es:

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N \left(\sum_{t=1}^T \hat{v}_{it} \right)^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2} - 1 \right]^2 \quad (37)$$

- Bajo la hipótesis nula (37) se distribuye como una χ_1^2 siempre y cuando los errores tengan **DISTRIBUCION NORMAL**.
- Como (36) no hace ningún supuesto sobre la distribución de los errores compuestos es preferible a (37).

Agenda

- 1 Introducción
- 2 Teoría Asintótica
- 3 Modelos de Datos de Panel Lineales
- 4 Modelos de Efectos Fijos y Aleatorios**
 - Modelo de Efectos Aleatorios
 - Modelo de Efectos Fijos**
 - Transformación de Efectos Fijos ó Within Transformation
 - Modelo de Variables Binarias (LSDV)
 - Transformación de Diferencias Finitas
 - Transformación por Desviaciones Ortogonales
- 5 Two-Way Fixed Effects Model

Modelo de Efectos Fijos

- Consideremos nuevamente el **modelo de componentes no observados**

$$\begin{aligned}y_{it} &= x_{it}\beta + c_i + u_{it} \quad \forall i, t \\ y_i &= X_i\beta + c_iJ_T + u_i \quad \forall i\end{aligned}\tag{38}$$

donde:

$$y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} ; \quad X_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iT} \end{bmatrix} ; \quad J_T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

- Como vimos, el procedimiento de RE para estimar β es poner a c_i en el término de error bajo el supuesto de que c_i es ortogonal a X_i y luego tomar en cuenta la correlación serial del error compuesto usando GLS.

Modelo de Efectos Fijos

- En muchas aplicaciones, todo el punto de trabajar con datos de panel es permitir que c_i y X_i estén arbitrariamente correlacionados.
- El modelo de **efectos fijos** asume que $\text{cov}(c_i, x_{it}) \neq 0$ y por lo tanto $E(v_{it}|X_i) \neq 0$. En otras palabras, una o más variables explicativas están correlacionadas con el error compuesto y FGLS dará estimaciones sesgadas e inconsistentes.
- Para obtener estimaciones consistentes, el **modelo de efectos fijos** asume **exogeneidad estricta de las variables explicativas condicionadas en c_i** .
- **Supuesto FE.1:** $E(u_{it}|X_i, c_i) = 0, \quad t = 1, 2, \dots, T$
- Note que este supuesto es exactamente el mismo que **RE.1(a)**.
- La diferencia fundamental con RE es que **no asumimos RE.1(b)**. Esto es, el análisis de efectos fijos permite que $E(c_i|X_i)$ sea una función de X_i .

Modelo de Efectos Fijos

- Relajando el supuesto RE.1(b) podemos estimar en forma consistente efectos parciales en presencia de variables omitidas constantes en el tiempo.
- En este último sentido, el análisis de FE es más robusto que el de RE.
- Sin embargo, esta mayor robustez tiene un precio.
- No podemos incluir en x_{it} factores constantes en el tiempo.
- La idea detrás de la estimación de β bajo el supuesto FE.1 es transformar (38) para eliminar el efecto no observable c_i .
- Existen varias transformaciones que logran eliminar c_i .
- Nosotros trabajaremos con tres transformaciones:
 - (a) Efectos fijos (FE ó within transformation)
 - (b) Diferencias finitas
 - (c) Desviaciones ortogonales (forward orthogonal deviations)

Modelo de Efectos Fijos

- Within Transformation
- La transformación de FE se obtiene promediando la ecuación (38) sobre $t = 1, 2, \dots, T$ para obtener la ecuación de corte transversal:

$$\bar{y}_i = \bar{x}_i \beta + c_i + \bar{u}_i \quad (39)$$

donde $\bar{y}_i = (1/T) \sum_{t=1}^T y_{it}$, $\bar{x}_i = (1/T) \sum_{t=1}^T x_{it}$ y $\bar{u}_i = (1/T) \sum_{t=1}^T u_{it}$.

- Restando miembro a miembro (39) de (38) se obtiene

$$\begin{aligned} y_{it} - \bar{y}_i &= (x_{it} - \bar{x}_i) \beta + u_{it} - \bar{u}_i \\ \ddot{y}_{it} &= \ddot{x}_{it} \beta + \ddot{u}_{it} \end{aligned} \quad (40)$$

donde $\ddot{y}_{it} = y_{it} - \bar{y}_i$, $\ddot{x}_{it} = x_{it} - \bar{x}_i$ y $\ddot{u}_{it} = u_{it} - \bar{u}_i$.

- Con c_i fuera de la ecuación es lógico pensar en estimar (40) por POLS.

Modelo de Efectos Fijos

- Recordemos que para obtener estimadores consistentes por POLS necesitamos que se cumplan los **supuestos 1 y 2**. Esto es:

$$E(\ddot{x}_{it}' \ddot{u}_{it}) = 0, \quad t = 1, 2, \dots, T \quad (41)$$

- Para cada t , (41) puede escribirse como:

$$E[(x_{it} - \bar{x}_i)'(u_{it} - \bar{u}_i)]$$

- Bajo el **supuesto FE.1** de exogeneidad estricta (41) se cumple.
- Por lo tanto, POLS puede aplicarse para obtener estimaciones consistentes.
- Note que el supuesto de exogeneidad estricta no puede relajarse a algo como exogeneidad contemporánea porque este último supuesto no garantiza que se cumpla (41).
- El **estimador de efectos fijos**, denotado por $\hat{\beta}^{FE}$ es el estimador POLS de la regresión de \ddot{y}_{it} sobre $\ddot{x}_{it} \forall i, t$.

Modelo de Efectos Fijos

- Entonces el estimador de efectos fijos de β , $\hat{\beta}^{FE}$ es:

$$\begin{aligned}\hat{\beta}^{FE} &= \left(\sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^N \ddot{X}_i' \ddot{y}_i \right) \\ &= \left(\sum_{i=1}^N X_i' Q_T X_i \right)^{-1} \left(\sum_{i=1}^N X_i' Q_T y_i \right) \\ &= \left(\ddot{X}' \ddot{X} \right)^{-1} \ddot{X}' \ddot{y} \\ &= [X'(I_N \otimes Q_T)X]^{-1} X'(I_N \otimes Q_T)y\end{aligned}\tag{42}$$

donde $Q_T = I_T - J_T(J_T' J_T)^{-1} J_T'$ es la matriz **time-demeaning** que es una matriz simétrica e idempotente de rango $T - 1$.

- Note que $Q_T \times J_T = 0$; $Q_T \times y_i = \ddot{y}_i$ y $Q_T \times X_i = \ddot{X}_i$.

Modelo de Efectos Fijos

- Para que el estimador de FE se comporte bien asintóticamente necesitamos la condición de rango estándar:
- **Supuesto FE.2:** $\text{rango} \left[\sum_{t=1}^T E(\ddot{x}_{it}' \ddot{x}_{it}) \right] = \text{rango} \left[E(\ddot{X}_i' \ddot{X}_i) \right] = K$
- Note que si x_{it} contiene algún elemento que no varía en el tiempo para cualquier i , entonces el elemento correspondiente en \ddot{x}_{it} es idénticamente igual a cero.
- Como \ddot{X}_i contendría una columna de ceros, el **supuesto FE.2** no podría ser verdadero.
- Esto muestra explícitamente porque las variables constantes en el tiempo no están permitidas en el análisis.
- El estimador de efectos fijos (42) recibe usualmente el nombre de **within estimator** porque utiliza la variación temporal dentro de cada corte transversal.

Modelo de Efectos Fijos

- Existe un segundo estimador de β conocido como el **between estimator**.
- El between estimator consiste en aplicar OLS a la ecuación promediada en el tiempo (39): $\bar{y}_i = \bar{x}_i\beta + c_i + \bar{u}_i$
- Este estimador **NO** es consistente bajo el supuesto FE.1 porque $E(\bar{x}_i'c_i) \neq 0$.
- Para obtener un estimador consistente en (39) necesitamos asumir **RE.1** y la **condición de rango estándar**.

Modelo de Efectos Fijos

- Inferencia Asintótica en FE
- El siguiente supuesto asegura que el estimador de FE es el más eficiente:
- **Supuesto FE.3:** $E(u_i u_i' | X_i, c_i) = \sigma_u^2 I_T$.
- El **supuesto FE.3** es idéntico a **RE.3(a)**.
- Como $E(u_i | X_i, c_i) = 0$ por FE.1, el supuesto FE.3 es igual a decir que $\text{Var}(u_i | X_i, c_i) = \sigma_u^2 I_T$.
- El supuesto FE.3 junto con el supuesto FE.1 aseguran que la matriz de varianzas y covarianzas marginal del error compuesto tiene la estructura que vimos para RE, pero sin el supuesto RE.3(b).
- Este resultado que es importante para RE no tiene ninguna importancia para hacer inferencia bajo FE.

► Back

Modelo de Efectos Fijos

- Considere la siguiente ecuación

$$\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it} \quad (43)$$

- Para que POLS aplicado a (43) resulte eficiente necesitamos que los errores sean homocedásticos y que no estén serialmente correlacionados en el tiempo.
- La varianza de \ddot{u}_{it} puede calcularse como:

$$\begin{aligned} E(\ddot{u}_{it}^2) &= E[(u_{it} - \bar{u}_i)^2] = E(u_{it}^2) + E(\bar{u}_i^2) - 2E(u_{it}\bar{u}_i) \\ &= \sigma_u^2 + \sigma_u^2/T - 2\sigma_u^2/T = \sigma_u^2(1 - 1/T) \end{aligned}$$

lo que verifica la homocedasticidad.

- La covarianza entre \ddot{u}_{it} y \ddot{u}_{is} es:

$$\begin{aligned} E(\ddot{u}_{it}\ddot{u}_{is}) &= E[(u_{it} - \bar{u}_i)(u_{is} - \bar{u}_i)] \\ &= E(u_{it}u_{is}) + E(\bar{u}_i^2) - E(u_{it}\bar{u}_i) - E(u_{is}\bar{u}_i) \\ &= 0 + \sigma_u^2/T - \sigma_u^2/T - \sigma_u^2/T = -\sigma_u^2/T \end{aligned}$$

Modelo de Efectos Fijos

- Combinando las dos expresiones anteriores tenemos:

$$\text{Corr}(\ddot{u}_{it}, \ddot{u}_{is}) = -1/(T - 1)$$

Lo que muestra que los errores transformados tienen correlación serial negativa en el tiempo. Para encontrar la varianza asintótica del estimador de FE escribamos (42) de la siguiente manera:

$$\begin{aligned}\hat{\beta}^{FE} &= \left(\sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^N \ddot{X}_i' \ddot{y}_i \right) \\ &= \beta + \left(\sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^N \ddot{X}_i' u_i \right) \Rightarrow \\ \sqrt{N}(\hat{\beta}^{FE} - \beta) &= \left(N^{-1} \sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(N^{-1/2} \sum_{i=1}^N \ddot{X}_i' u_i \right)\end{aligned}\tag{44}$$

Modelo de Efectos Fijos

- La ecuación anterior surge de utilizar la siguiente relación:

$$\ddot{X}_i' \ddot{u}_i = (Q_T X_i)' Q_T u_i = X_i' Q_T u_i = \ddot{X}_i' u_i$$

- Por el supuesto FE.3: $E(u_i u_i' | X_i, c_i) = \sigma_u^2 I_T$.
- Por lo tanto:

$$\sqrt{N}(\hat{\beta}^{FE} - \beta) \xrightarrow{d} \text{Normal} \left(0, \sigma_u^2 [E(\ddot{X}_i' \ddot{X}_i)]^{-1} \right)$$

- Y además: $\text{Avar}(\hat{\beta}^{FE}) = \sigma_u^2 [E(\ddot{X}_i' \ddot{X}_i)]^{-1} / N$
- Dado un estimador consistente de σ_u^2 , la varianza asintótica puede ser estimada reemplazando la esperanza por su análogo muestral.

$$\widehat{\text{Avar}}(\hat{\beta}^{FE}) = \hat{\sigma}_u^2 \left(\sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} = \hat{\sigma}_u^2 \left(\sum_{i=1}^N \sum_{t=1}^T \ddot{x}_{it}' \ddot{x}_{it} \right)^{-1} \quad (45)$$

Modelo de Efectos Fijos

- Los errores estándar asintóticos se obtienen con la raíz cuadrada de los elementos de la diagonal principal de (45).
- El único punto a tener en cuenta es la estimación de σ_u^2 .
- Note que sumando sobre t , $E(\ddot{u}_{it}^2)$ obtenemos $(T - 1)\sigma_u^2$. Por lo tanto:

$$\frac{1}{T - 1} \sum_{t=1}^T E(\ddot{u}_{it}^2) = \sigma_u^2 \quad (46)$$

- Definamos los residuos de FE como:

$$\hat{\ddot{u}}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\hat{\beta}^{FE}, \quad \forall i, t$$

- Un estimador consistente de σ_u^2 es entonces:

$$\hat{\sigma}_u^2 = \frac{1}{N(T - 1) - K} \sum_{i=1}^N \sum_{t=1}^T \hat{\ddot{u}}_{it}^2 \quad (47)$$

Modelo de Efectos Fijos

- Piense que uno podría haber conseguido un estimador para σ_u^2 aplicando el principio de analogía en (46) y reemplazar la esperanza por su análogo muestral.
- Un punto a tener en cuenta es que el denominador de (47) NO son los grados de libertad que uno obtendría de aplicar POLS a la ecuación transformada por FE (43).
- La estimación de la varianza de los errores en (43) sería $RSS/(NT - K)$.
- La diferencia entre esta última estimación y (47) puede ser grande si T es chico.
- En general los errores estándar reportados directamente de (43) tienden a ser pequeños comparados con los verdaderos.

Modelo de Efectos Fijos

- Bajo los **supuestos FE.1-FE.3** restricciones múltiples en los coeficientes pueden ser contrastadas utilizando la fórmula estándar del test de Wald:

$$F = \frac{(RSS_r - RSS_u)/Q}{RSS_u/[N(T-1) - K]} \xrightarrow{d} F_{Q, N(T-1)-K}$$

Modelo de Efectos Fijos

- El Modelo de Variables Binarias (LSDV)
- El enfoque tradicional de FE es ver a c_i como parámetros a ser estimados.
- Si cambiamos el supuesto FE.2 a su versión de muestra finita: $\text{rango} \ddot{X}' \ddot{X} = K$, el modelo satisface todos los supuestos de Gauss-Markov.
- Para estimar el modelo se definen N variables binarias, una para cada observación de corte transversal.
- Luego se estima por POLS una regresión de y_{it} sobre las variables binarias, x_{it} con $t = 1, 2, \dots, T$; $j = 1, 2, \dots, N$.
- Los coeficientes que acompañan a las variables binarias son las estimaciones de los c_i .
- El estimador obtenido de esta última regresión es exactamente igual al estimador de FE. Debido a esto, el estimador de FE recibe el nombre de **estimador de variables binarias**.

Modelo de Efectos Fijos

- Considere el siguiente modelo,

$$y = X\beta + C\gamma + u = X\beta + (I_N \otimes J_T)\gamma + u \quad (48)$$

donde γ es $N \times 1$ y es el vector de coeficientes que acompañan a las variables binarias.

- Re-escribiendo (48) como

$$y = [X|(I_N \otimes J_T)] \begin{pmatrix} \beta \\ \gamma \end{pmatrix} + u \quad (49)$$

- Obtenemos,

$$\hat{\beta}^{LSDV} = [X'(I_N \otimes Q_T)X]^{-1} X'(I_N \otimes Q_T)y \quad (50)$$

- Los residuos del modelo LSDV son exactamente iguales a los de (43).
- Una ventaja del modelo LSDV es que produce el estimador correcto de la varianza de los errores porque usa como grados de libertad $NT - N - K = N(T - 1) - K$.

Modelo de Efectos Fijos

- Un problema con el enfoque LSDV es que los estimadores de c_i son insesgados pero no son consistentes.
- El estimador de FE es consistente y asintóticamente normal si se cumplen FE.1 y FE.2.
- Si no se cumple FE.3, entonces (45) nos dará un estimador incorrecto de la matriz de varianzas y covarianzas de los estimadores.
- Si no se cumple FE.3, entonces debemos reemplazar (45) por una estimación robusta.
- Aplicando los resultados ya vistos, podemos utilizar la ecuación (13) reemplazando los residuos por los estimados por FE.

$$\hat{V} = \left(\sum_{j=1}^N \ddot{X}_j' \ddot{X}_j \right)^{-1} \left(\sum_{j=1}^N \ddot{X}_j' \hat{u}_j \hat{u}_j' \ddot{X}_j \right) \left(\sum_{j=1}^N \ddot{X}_j' \ddot{X}_j \right)^{-1} \quad (51)$$

- Los errores estándar de los estimadores de FE se obtienen de la raíz cuadrada de los elementos de la diagonal principal de (51).

Modelo de Efectos Fijos

- En lugar de calcular una matriz de varianzas y covarianzas robusta para el estimador FE, se podría relajar el supuesto FE.3 para permitir una matriz no restringida general y aplicar GLS.
- Supuesto FEGLS.3: $E(u_i u_i' | X_i, c_i) = \Lambda$, una matriz $T \times T$ definida positiva.
- Bajo FEGLS.3, $E(\ddot{u}_i \ddot{u}_i' | \ddot{X}_i) = E(\ddot{u}_i \ddot{u}_i')$ y usando el hecho de que $\ddot{u}_i = Q_T u_i$, tenemos,
- $E(\ddot{u}_i \ddot{u}_i') = Q_T \Lambda Q_T$, que tiene rango $T - 1$.
- Esto es un problema porque no podemos invertir esta matriz para obtener los estimadores GLS.
- Dos soluciones:
 - (i) trabajar con la inversa generalizada.
 - (ii) eliminar un período temporal.

Modelo de Efectos Fijos

- Supongamos que eliminamos el período T . Entonces tenemos las siguientes ecuaciones:

$$\begin{aligned}\ddot{y}_{i1} &= \ddot{x}_{i1}\beta + \ddot{u}_{i1} \\ \vdots &= \vdots \quad \quad \quad \vdots \\ \ddot{y}_{iT-1} &= \ddot{x}_{iT-1}\beta + \ddot{u}_{iT-1}\end{aligned}$$

- Podemos re-escribir este sistema como si fuera (42) con la única diferencia que ahora los vectores y matrices tienen dimensión $(T - 1)$.
- Definamos la matriz $(T - 1) \times (T - 1)$: $\Omega = E(\ddot{u}_i \ddot{u}_i')$
- Para estimar Ω , tenemos que estimar β por FE en un primer paso.
- Después hay que eliminar el período T y construir los $(T - 1) \times 1$ residuos

$$\hat{\hat{u}}_i = \ddot{y}_i - \ddot{X}_i \hat{\beta}^{FE}, \quad i = 1, 2, \dots, N$$

Modelo de Efectos Fijos

- Un estimador consistente de Ω es:

$$\hat{\Omega} = N^{-1} \sum_{i=1}^N \hat{u}_i \hat{u}_i'$$

- El estimador de FE por GLS se define como:

$$\hat{\beta}^{FEGLS} = \left(\frac{1}{N} \sum_{j=1}^N \ddot{X}_j' \hat{\Omega}^{-1} \ddot{X}_j \right)^{-1} \left(\frac{1}{N} \sum_{j=1}^N \ddot{X}_j' \hat{\Omega}^{-1} \ddot{y}_j \right) \quad (52)$$

donde las variables están definidas sin el último período temporal.

- Para obtener consistencia necesitamos una nueva condición de rango:
- Supuesto FEGLS.2:** $\text{rango} E(\ddot{X}_j' \Omega^{-1} \ddot{X}_j) = K$.
- Bajo FE.1 y FEGLS.2 el estimador de FEGLS es consistente.

Modelo de Efectos Fijos

- Adicionando el supuesto FEGLS.3, la matriz de varianzas y covarianzas asintótica se estima por:

$$\widehat{\text{Avar}}(\hat{\beta}^{FEGLS}) = \left(\sum_{j=1}^N \ddot{X}_i' \hat{\Omega}^{-1} \ddot{X}_i \right)^{-1}$$

Modelo de Efectos Fijos

- Aparte de la **within transformation** existen otras transformaciones para eliminar la heterogeneidad no observada. Una de las más utilizadas es la **transformación de diferencias finitas (FD)**.
- Considere el modelo de componentes no observados escrito para los períodos t y $t - 1$,

$$y_{it} = \alpha_i + x_{it}\beta + u_{it} \quad (53)$$

$$y_{it-1} = \alpha_i + x_{it-1}\beta + u_{it-1} \quad (54)$$

- Restando miembro a miembro,

$$y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + (u_{it} - u_{it-1}) \quad (55)$$

$$\Delta y_{it} = \Delta x_{it}\beta + \Delta u_{it} \quad (56)$$

- Como ocurría con la transformación de FE, FD también elimina el efecto individual

Modelo de Efectos Fijos

- El estimador de diferencias finitas de β , $\hat{\beta}^{FD}$, es:

$$\begin{aligned}\hat{\beta}^{FD} &= \left(\sum_{i=1}^N \Delta X_i' \Delta X_i \right)^{-1} \left(\sum_{i=1}^N \Delta X_i' \Delta y_i \right) \\ &= \left(\sum_{i=1}^N (DX_i)' DX_i \right)^{-1} \left(\sum_{i=1}^N (DX_i)' Dy_i \right) \\ &= \left(\sum_{i=1}^N X_i' D' DX_i \right)^{-1} \left(\sum_{i=1}^N X_i' D' Dy_i \right)\end{aligned}\tag{57}$$

donde

$$D = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & -1 & 1 \end{bmatrix}\tag{58}$$

Modelo de Efectos Fijos

- Note que para que este estimador sea consistente necesitamos que se cumpla que:
Supuesto FD.1 : $E(u_{it}|X_i, c_i) = 0, \quad t = 1, 2, \dots, T$
- El **supuesto FD.1**, es igual a **FE.1** y a **RE.1(a)**.
- Bajo **supuesto FD.1**, POLS en (56) será consistente porque $E(\Delta x'_{it} \Delta u_{it}) = 0, \quad t = 2, \dots, T$, respectivamente.
- Recuerde que además de este supuesto necesitamos una condición de rango estándar: **supuesto FD.2**: $\text{rango} E(\Delta X'_i \Delta X_i) = K$.
- Una de las razones para preferir FD sobre FE es que es fácil de calcular usando un paquete estadístico común.
- Lo único que debemos tener en cuenta es que las observaciones correspondientes a los períodos $1, T + 1, 2T + 1, \dots, (N - 1)T + 1$ deben considerarse como no disponibles.

Modelo de Efectos Fijos

- Sin embargo, bajo los **supuestos FE.1-FE.3**, el estimador de FE es el más eficiente dentro de la clase de estimadores que utilizan el supuesto de exogeneidad estricta.
- Una consecuencia de este último punto es que el estimador de **FD** debe ser **menos eficiente si se cumple FE.3**. ► FE.3
- Si se cumple **FE.3**, entonces: $E(Du_i u_i' D' | X_i, c_i) = DE(u_i u_i' | X_i, c_i) D' = \sigma_u^2 DD'$
- Lo que muestra que los errores estarán correlacionados para períodos adyacentes.
- En este caso, argumentos estándares de GLS nos darán el siguiente estimador óptimo:

$$\hat{\beta}^{FDGLS} = \left(\sum_{i=1}^N X_i' D' (DD')^{-1} D X_i \right)^{-1} \left(\sum_{i=1}^N X_i' D' (DD')^{-1} D y_i \right) \quad (59)$$

Modelo de Efectos Fijos

- Note que en este caso $GLS \equiv FGLS$ porque DD' es conocida.
- Otro punto interesante es que, la matriz idempotente $D'(DD')^{-1}D$ también puede escribirse como:

$$D'(DD')^{-1}D \equiv I_T - J_T(J_T'J_T)^{-1}J_T' = Q_T$$

donde Q_T es la matriz **time demeaning** que vimos antes.

- Esto muestra que: $\hat{\beta}^{FDGLS} = \hat{\beta}^{FE}$
- Si no se cumple **FE.3** entonces uno puede asumir que la primera diferencia de los errores idiosincráticos no tienen correlación serial.
- **Supuesto FD.3:** $E(e_i e_i' | X_i, c_i) = \sigma_e^2 I_{T-1}$, donde e_i es el vector que contiene a $e_{it} = \Delta u_{it}$, $t = 2, \dots, T$ (ó $e_i = Du_i$).

Modelo de Efectos Fijos

- Bajo el supuesto FD.3, podemos escribir los errores idiosincráticos como:

$$u_{it} = u_{it-1} + e_{it}$$

- Tal que, ausencia de correlación serial en e_{it} implica que u_{it} sigue un paseo al azar (i.e. tiene una dependencia serial muy fuerte).
- Es decir que el supuesto FD.3 representa el otro extremo de FE.3
- Bajo los supuestos FD.1-FD.3, el estimador de FD es el más eficiente dentro de la clase de estimadores que cumplen FE.1.
- Utilizando los resultados de POLS en (57) tenemos que:

$$\widehat{\text{Avar}}(\widehat{\beta}^{FD}) = \hat{\sigma}_e^2 \left(\sum_{i=1}^N X_i' D' D X_i \right)^{-1}$$

donde $\hat{\sigma}_e^2$ es un estimador consistente de σ_e^2 .

Modelo de Efectos Fijos

- El estimador más simple se obtiene calculando los residuos

$$\hat{e}_{it} = y_{it} - x_{it}\hat{\beta}^{FD}$$

- Y luego estimando σ_e^2 como:

$$\hat{\sigma}_e^2 = \frac{1}{N(T-1) - K} \sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2$$

- Si el supuesto FD.3 no se cumple, entonces debemos utilizar una matriz de varianzas y covarianzas robusta.
- Usando (13) tenemos:

$$\hat{V} = \left(\sum_{i=1}^N X_i' D' D X_i \right)^{-1} \left(\sum_{i=1}^N X_i' D' \hat{e}_i \hat{e}_i' D X_i \right) \left(\sum_{i=1}^N X_i' D' D X_i \right)^{-1} \quad (60)$$

Modelo de Efectos Fijos

- Transformación de Helmert
- Otra forma de eliminar el efecto individual en (38) es mediante la transformación de Helmert ó desviaciones ortogonales hacia adelante:

$$y_{it}^* = \sqrt{\frac{T-t}{T-t+1}} \left[y_{it} - \frac{1}{T-t} (y_{it+1} + \dots + y_{iT}) \right], \quad t = 1, 2, \dots, T-1$$

- El modelo transformado es entonces:

$$y_{it}^* = x_{it}^* \beta + u_{it}^* \quad (61)$$

- Tal como ocurría con FD y FE, la transformación de Helmert elimina el efecto individual.
- El estimador de desviaciones ortogonales es el estimador POLS en (61).

Modelo de Efectos Fijos

- Definamos la matriz H_T de dimensión $(T - 1) \times T$: $H_T = (DD')^{-1/2}D$.
- Eligiendo $(DD')^{-1/2}$ como la matriz triangular superior que surge de la factorización de Cholesky, tenemos:

$$H_T = \text{diag}[(T - 1)/T, \dots, 1/2]^{-1/2}H^+$$

con

$$H^+ \equiv \begin{bmatrix} 1 & (1 - T)^{-1} & (1 - T)^{-1} & \dots & (1 - T)^{-1} & (1 - T)^{-1} & (1 - T)^{-1} \\ 0 & 1 & (2 - T)^{-1} & \dots & (2 - T)^{-1} & (2 - T)^{-1} & (2 - T)^{-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & \dots & 0 & 1 & -1 \end{bmatrix}$$

Modelo de Efectos Fijos

- Usando esta matriz H_T el modelo (61) puede escribirse como:

$$H_T y_i = H_T x_i \beta + H_T u_i$$

- El estimador HT es entonces:

$$\begin{aligned}\hat{\beta}^{HT} &= \left(\sum_{i=1}^N X_i' H_T' H_T X_i \right)^{-1} \left(\sum_{i=1}^N X_i' H_T' H_T y_i \right) \\ &= \beta + \left(\sum_{i=1}^N X_i' H_T' H_T X_i \right)^{-1} \left(\sum_{i=1}^N X_i' H_T' H_T u_i \right)\end{aligned}\tag{62}$$

- Bajo que supuestos este estimador es consistente?
- Claramente como $H_T = (DD')^{-1/2}D$, entonces $H_T' H_T = Q_T$. Y por lo tanto

$$\hat{\beta}^{FE} \equiv \hat{\beta}^{FDGLS} \equiv \hat{\beta}^{HT}$$

Modelo de Efectos Fijos

- El último punto implica que necesitamos:
- **Supuesto HT.1:** $E(u_{it}|X_i, c_i) = 0, \quad t = 1, 2, \dots, T$
- El **supuesto HT.1** es igual a **FE.1**, a **FD.1** y a **RE.1(a)**.
- **Supuesto HT.2:** $\text{rango}[E(X_i' H_T' H_T X_i)] = K$
- Bajo los **supuestos HT.1 y HT.2**, el estimador $\hat{\beta}^{HT}$ es consistente.
- Note que como $H_T = (DD')^{-1/2}D$, entonces $H_T H_T' = I_{T-1}$, y por lo tanto si se cumple **FE.3**:

$$E(u_i^* u_i^{*'} | X_i, c_i) = H_T E(u_i u_i' | X_i, c_i) H_T' = \sigma_u^2 I_{T-1}.$$

- Queda claro de (62) que

$$\begin{aligned} \hat{\beta}^{HT} &= \beta + \left(\sum_{i=1}^N X_i' H_T' H_T X_i \right)^{-1} \left(\sum_{i=1}^N X_i' H_T' H_T u_i \right) \\ &= \beta + \left(\sum_{i=1}^N X_i' Q_T X_i \right)^{-1} \left(\sum_{i=1}^N X_i' Q_T u_i \right) = \hat{\beta}^{FE} \end{aligned} \quad (63)$$

Modelo de Efectos Fijos

- Por lo tanto, bajo HT.1-HT.2 y FE.3,

$$\sqrt{N}(\hat{\beta}^{HT} - \beta) \xrightarrow{d} \text{Normal}(0, \sigma_u^2 [E(X_i' H_T' H_T X_i)]^{-1})$$

- Y además: $\text{Avar}(\hat{\beta}^{HT}) = \sigma_u^2 [E(X_i' H_T' H_T X_i)]^{-1} / N$.
- Dado un estimador consistente de σ_u^2 , la varianza asintótica puede ser estimada reemplazando la esperanza por su análogo muestral.

$$\widehat{\text{Avar}}(\hat{\beta}^{HT}) = \hat{\sigma}_u^2 \left[\sum_{i=1}^N X_i' H_T' H_T X_i \right]^{-1} \quad (64)$$

- Los errores estándar asintóticos se obtienen con la raíz cuadrada de los elementos de la diagonal principal de (64).
- Al igual que ocurría con la transformación de FE, el único punto a tener en cuenta es la estimación de σ_u^2 .

Modelo de Efectos Fijos

- Un estimador consistente de σ_u^2 es:

$$\hat{\sigma}_u^2 = \frac{1}{N(T-1) - K} \sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^* \hat{u}_{it}^{*'} \quad (65)$$

donde, $\hat{u}_{it}^* = y_{it}^* - x_{it}^* \hat{\beta}^{HT}$.

- Si FE.3 no se satisface, entonces debemos reemplazar (64) por una estimación robusta.
- Aplicando los resultados ya vistos, podemos utilizar la ecuación (13) reemplazando los residuos por los estimados por HT.

$$\hat{V} = \left(\sum_{i=1}^N x_i^{*'} x_i^* \right)^{-1} \left(\sum_{i=1}^N x_i^{*'} \hat{u}_i^* \hat{u}_i^{*'} x_i^* \right) \left(\sum_{i=1}^N x_i^{*'} x_i^* \right)^{-1} \quad (66)$$

- Los errores estándar de los estimadores de HT se obtienen de la raíz cuadrada de los elementos de la diagonal principal de (66).

Modelo de Efectos Fijos

- Relación entre FE y RE
- Escribamos la matriz de varianzas y covarianzas con la estructura de RE:

$$\begin{aligned}\Omega &= \sigma_u^2 I_T + \sigma_c^2 J_T J_T' = \sigma_u^2 I_T + T \sigma_c^2 J_T (J_T' J_T)^{-1} J_T' \\ &= \sigma_u^2 I_T + T \sigma_c^2 P_T = (\sigma_u^2 + T \sigma_c^2)(P_T + \eta Q_T)\end{aligned}$$

donde $P_T \equiv I_T - Q_T = J_T(J_T' J_T)^{-1} J_T'$, y $\eta = \sigma_u^2 / (\sigma_u^2 + T \sigma_c^2)$.

- Note que de la definición de P_T tenemos las siguientes relaciones:
 - (i) $P_T + Q_T = I_T$
 - (ii) $P_T Q_T = 0$
 - (iii) $P_T P_T = P_T$
- Ahora definamos $S_T = P_T + \eta Q_T$. $S_T^{-1} = P_T + (1/\eta)Q_T$ y $S_T^{-1/2} = P_T + (1/\sqrt{\eta})Q_T$.

Modelo de Efectos Fijos

- Usando álgebra: $S_T^{-1/2} = (1 - \lambda)^{-1}[I_T - \lambda P_T]$, con $\lambda = 1 - \sqrt{\eta}$.
- Por lo tanto,

$$\begin{aligned}\Omega^{-1/2} &= (\sigma_u^2 + T\sigma_c^2)^{-1/2}(1 - \lambda)^{-1}[I_T - \lambda P_T] \\ &= (1/\sqrt{\sigma_u^2})[I_T - \lambda P_T]\end{aligned}$$

donde $\lambda = 1 - [\sigma_u^2/(\sigma_u^2 + T\sigma_c^2)]^{1/2}$.

- Asumamos por un momento que conocemos λ . Entonces, RE se obtiene con la ecuación transformada: $C_T y_i = C_T X_i \beta + C_T u_i$ con $C_T = [I_T - \lambda P_T]$.
- Escribamos la ecuación transformada como:

$$\check{y}_i = \check{X}_i \beta + \check{u}_i$$

- La varianza de \check{u}_i es $E(\check{u}_i \check{u}_i') = C_T \Omega C_T = \sigma_u^2 I_T$.

Modelo de Efectos Fijos

- Claramente el elemento t de \check{y}_i es

$$y_{it} - \lambda \bar{y}_i$$

- Por lo tanto RE es POLS en:

$$y_{it} - \lambda \bar{y}_i = (x_{it} - \lambda \bar{x}_i)\beta + (u_{it} - \lambda \bar{u}_i), \quad \forall i, t \quad (67)$$

- Los errores de esta ecuación son homocedásticos y no están serialmente correlacionados bajo el **supuesto RE.3**.
- FGLS se obtiene reemplazando λ con un estimador consistente.
- Si $\hat{\lambda}$ es un estimador consistente de λ , entonces:

$$\hat{\beta}^{RE} = \left(\sum_{i=1}^N \sum_{t=1}^T \check{x}'_{it} \check{x}_{it} \right)^{-1} \sum_{i=1}^N \sum_{t=1}^T \check{x}'_{it} \check{y}_{it} \quad (68)$$

Modelo de Efectos Fijos

- Los estadísticos t y F usuales son válidos asintóticamente bajo los **supuestos RE.1-RE.3**.
- La ecuación (68) muestra que el estimador de RE se puede obtener con lo que se denomina **quasi-time-demeaning**.
- En lugar de sacarle la media temporal, los RE le sacan una fracción $\hat{\lambda}$ de la media temporal a las variables.
- Si $\hat{\lambda}$ es cercano a 1, entonces **RE y FE tienden a acercarse**.
- Para ver cuando esto ocurre escribamos $\hat{\lambda}$ como:

$$\hat{\lambda} = 1 - \{1/[1 + T(\sigma_c^2/\sigma_u^2)]\}^{1/2}$$

- Por lo tanto: $\hat{\lambda} \longrightarrow 1$ **cuando** $T \longrightarrow \infty$ **ó** $(\sigma_c^2/\sigma_u^2) \longrightarrow \infty$.

Inferencia en Modelos de Panel

- Como la consideración fundamental para elegir entre FE y RE es el hecho de que los efectos no observables estén o no correlacionados con las variables explicativas, es importante tener un test que contraste este supuesto.
- Hausman (1978) propuso un test basado en las diferencias entre los estimadores de FE y RE.
- La hipótesis nula del test asume no correlación entre c_i y x_{it} por lo tanto ambos FE y RE son consistentes, pero FE es ineficiente.
- La hipótesis alternativa asume que hay correlación entre c_i y x_{it} por lo tanto FE es consistente, pero RE no.
- Por lo tanto bajo la nula, los dos estimadores no debieran diferir mucho.
- Denotemos por $\hat{\delta}^{RE}$ al vector de estimadores de RE sin los coeficientes que acompañan a variables constantes en el tiempo; y por $\hat{\delta}^{FE}$ a los correspondientes estimadores de FE.

Inferencia en Modelos de Panel

- Suponiendo que ambos vectores de estimadores tienen dimensión $M \times 1$, el estadístico del test de Hausman es:

$$H = (\hat{\delta}^{FE} - \hat{\delta}^{RE})' [\widehat{Var}(\delta)^{FE} - \widehat{Var}(\delta)^{RE}]^{-1} (\hat{\delta}^{FE} - \hat{\delta}^{RE}) \Rightarrow \chi_M^2 \quad (69)$$

- Si estamos interesados en un único parámetro podemos transformar el test de Hausman en un test t .
- Asumamos que estamos interesados en δ_1 . El test se vuelve:

$$t = \frac{(\hat{\delta}_1^{FE} - \hat{\delta}_1^{RE})}{\sqrt{[\widehat{Var}(\delta_1)^{FE} - \widehat{Var}(\delta_1)^{RE}]}} \Rightarrow N(0, 1) \quad (70)$$

Inferencia en Modelos de Panel

- **Remark 1:** Se mantienen el supuesto de exogeneidad estricta RE.1(a) bajo la nula y la alternativa.
- **Remark 2:** El test se implementa usualmente asumiendo que RE.3 se cumple bajo la nula.

Extensión del Modelo de Efectos Fijos

- Una generalización del modelo de componentes no observados incluye efectos fijos temporales.
- Esto es:

$$y_{it} = \alpha_i + \lambda_t + x_{it}\beta + u_{it}, \quad i \in \{1, \dots, N\}, \quad t \in \{1, \dots, T\}$$

- La inclusión de α_i controla por factores no observables específicos de las unidades pero constantes en el tiempo.
- La inclusión de λ_t controla por factores específicos en el tiempo pero constantes en el corte transversal.
- Ahora, además de calcular las medias temporales de cada variable, $\bar{y}_i = (1/T) \sum_{t=1}^T y_{it}$, necesitamos calcular las medias a través del corte transversal: $\bar{y}_t = (1/N) \sum_{i=1}^N y_{it}$ y las medias totales: $\bar{y} = (1/NT) \sum_{i=1}^N \sum_{t=1}^T y_{it}$

Extensión del Modelo de Efectos Fijos

- Para estimar este modelo considere la siguiente transformación:

$$y_{it} = \alpha_i + \lambda_t + x_{it}\beta + u_{it} \quad (71)$$

$$\bar{y}_i = \alpha_i + \bar{\lambda} + \bar{x}_i\beta + \bar{u}_i \quad (72)$$

$$\bar{y}_t = \bar{\alpha} + \lambda_t + \bar{x}_t\beta + \bar{u}_t \quad (73)$$

$$\bar{y} = \bar{\alpha} + \bar{\lambda} + \bar{x}\beta + \bar{u} \quad (74)$$

Extensión del Modelo de Efectos Fijos

- Restando miembro a miembro de la ecuación (71) la (72) y la (73) y sumando la (74) tenemos

$$\begin{aligned}y_{it} - \bar{y}_i - \bar{y}_t + \bar{y} &= (x_{it} - \bar{x}_i - \bar{x}_t + \bar{x})\beta + (u_{it} - \bar{u}_i - \bar{u}_t + \bar{u}) \\ \ddot{y}_{it} &= \ddot{x}_{it}\beta + \ddot{u}_{it}\end{aligned}\tag{75}$$

- El estimador de MCC de la ecuación (75) se conoce en la literatura como **two-way fixed effects estimator**.

Econometría de Datos de Panel

Maestrías en Economía y Econometría

Lecture 2

- 1 Modelos Dinámicos
 - Introducción a Modelos de Datos de Panel Dinámicos
 - El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
 - Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
 - El Estimador de Arellano-Bond
 - El Estimador de Blundell-Bond
 - Extensión: Regresores Exógenos
 - Contrastes de Validez de los Instrumentos
 - Datos con Persistencia

- 1 Modelos Dinámicos
 - Introducción a Modelos de Datos de Panel Dinámicos
 - El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
 - Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
 - El Estimador de Arellano-Bond
 - El Estimador de Blundell-Bond
 - Extensión: Regresores Exógenos
 - Contrastes de Validez de los Instrumentos
 - Datos con Persistencia

Modelos Dinámicos

- En los modelos que hemos visto hasta ahora se asumió que las variables explicativas eran estrictamente exógenas (en el caso de FE condicional al efecto no observable).
- En general, FE y RE son inconsistentes si existe correlación entre el error idiosincrático y alguna variable explicativa en algún período.
- Necesitamos una forma de estimación consistente, con $N \rightarrow \infty$ y T fijo, cuando las variables explicativas no son estrictamente exógenas.
- Esto es lo que ocurre cuando tenemos modelos dinámicos (i.e. la variable dependiente aparece como regresor rezagada).
- El modelo que vamos a analizar es el mismo que en FE:

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1, 2, \dots, T \quad (1)$$

Modelos Dinámicos

- Pero además de permitir que c_i y x_{it} estén arbitrariamente correlacionadas, ahora también permitimos que u_{it} esté correlacionada con valores futuros de las variables explicativas, $(x_{it+1}, x_{it+2}, \dots, x_{it+T})$.
- Ejemplo: $AR(1)$

$$y_{it} = y_{it-1}\beta + c_i + u_{it}, \quad t = 1, 2, \dots, T$$

- En este ejemplo $x_{it} = y_{it-1}$ por lo tanto u_{it} va a estar correlacionado con $x_{it+1} = y_{it}$.
- Para resolver este problema necesitamos una nueva condición de exogeneidad: **exogeneidad secuencial** (Chamberlain, 1992)
- Decimos que las variables explicativas son **secuencialmente exógenas condicionadas en el efecto no observable** cuando se cumple que:

$$E(u_{it} | x_{it}, x_{it-1}, \dots, x_{i1}, c_i) = 0, \quad t = 1, 2, \dots, T. \quad (2)$$

Modelos Dinámicos

- Usando el modelo (1) esta última condición es equivalente a:

$$E(y_{it}|x_{it}, x_{it-1}, \dots, x_{i1}, c_i) = E(y_{it}|x_{it}, c_i) = x_{it}\beta + c_i.$$

- La primera igualdad es la que le da el sentido a la condición: **exogeneidad secuencial** implica que después de haber controlado por x_{it} y c_i , ningún valor pasado de x_{it} afecta el valor esperado de y_{it} .
- Si estimamos por FE cuando el supuesto de exogeneidad estricta no se cumple obtendremos estimadores inconsistentes.
- Considere un modelo de panel para variables observadas a través de un corte transversal y en el tiempo, siendo β el parámetro que estamos interesados en estimar y con c_i siendo la **heterogeneidad no observada**:

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad i = 1, \dots, N \quad t = 1, \dots, T$$

Introducción a los Modelos de Panel Dinámicos

- Este modelo se puede re-escribir *stacking* las observaciones de series temporales como:

$$y_i = X_i\beta + c_i + u_i$$

- Los métodos de panel tradicionales para estimar β , como los modelos de **fixed-effect** o **random-effect**, transforman el modelo de forma tal de erradicar los problemas que causan la presencia de c_i . Esto lo hacen poniendo a c_i como parte del error y estimando β por FGLS o eliminándolo a través de primeras diferencias o con la *within transformation*.
- Estos métodos descansan en algún supuesto de exogeneidad para alcanzar la consistencia en términos de una teoría asintótica con T fijo. Considere, por ejemplo, el estimador de efectos fijos de β . Haciendo que las variables expresadas como desviaciones de sus medias temporales se denoten por $\tilde{\cdot}$, el estimador de efectos fijos de β , $\hat{\beta}^{FE}$ es:

Introducción a los Modelos de Panel Dinámicos



$$\begin{aligned}\widehat{\beta}^{FE} &= \beta + \left(\sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^N \ddot{X}_i' \ddot{u}_i \right) \\ &= \beta + \left(\sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\sum_{i=1}^N \ddot{X}_i' u_i \right) \\ &= \beta + \left(\frac{1}{N} \sum_{i=1}^N \ddot{X}_i' \ddot{X}_i \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^N \ddot{X}_i' u_i \right)\end{aligned}$$

- Para chequear las propiedades asintóticas (con T fijo) de $\widehat{\beta}^{FE}$, asumamos que tenemos una muestra aleatoria de las observaciones de corte transversal. Bajo la muestra aleatoria,

$$p \lim \left(\sum_{i=1}^N \ddot{X}_i' u_i \right) = E \left(\ddot{X}_i' u_i \right)$$

Introducción a los Modelos de Panel Dinámicos

- La consistencia requiere que,

$$\forall i \in \mathcal{N} : E \left(\ddot{X}_i' u_i \right) = 0_k$$

- Esta condición puede asegurarse claramente por el supuesto usual de **exogeneidad estricta**:

$$\forall i \in \mathcal{N}, \forall t \in \mathcal{T} : E(u_{it} | c_i, X_i) = E(u_{it} | c_i, x_{i1}, x_{i2}, \dots, x_{iT}) = 0$$

- Nosotros vamos a considerar un modelo de datos de panel **dinámico** cuando se cumpla la siguiente condición, más débil, de exogeneidad, **exogeneidad secuencial**:

$$\forall i \in \mathcal{N} : E(u_{it} | c_i, x_{i1}, \dots, x_{it-1}, x_{it}) = 0$$

Introducción a los Modelos de Panel Dinámicos

- Bajo exogeneidad secuencial, la condición de consistencia para el estimador de efectos fijos no puede garantizarse que se cumpla. Para ver porqué, calculemos $E(\ddot{x}_{it} u_{it})$:

$$\begin{aligned} E(\ddot{x}_{it} u_{it}) &= E \left[\left(x_{it} - \frac{1}{T} \sum_{t=1}^T x_{it} \right) u_{it} \right] \\ &= E[x_{it} u_{it}] - \frac{1}{T} \sum_{j=1}^T E[x_{ij} u_{it}] \end{aligned}$$

- Por la ley de expectativas iteradas y la condición de exogeneidad secuencial, tenemos

$$\begin{aligned} E[x_{it} u_{it}] &= E[x_{it} E[u_{it} | c_i, x_{i1}, \dots, x_{it-1}, x_{it}]] \\ &= 0 \end{aligned}$$

Introducción a los Modelos de Panel Dinámicos



$$E[x_{ij}u_{it}] = 0 \quad \forall j \leq t$$
$$\Rightarrow E(\ddot{x}_{it}u_{it}) = -\frac{1}{T} \sum_{j=t+1}^T E[x_{ij}u_{it}]$$

que no puede asumirse igual a cero. Por lo tanto, $\hat{\beta}^{FE}$ no es consistente en el contexto de modelos de panel dinámicos. La estimación consistente de β requiere de un nuevo método.

1 Modelos Dinámicos

- Introducción a Modelos de Datos de Panel Dinámicos
- El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
- Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
- El Estimador de Arellano-Bond
- El Estimador de Blundell-Bond
- Extensión: Regresores Exógenos
- Contrastes de Validez de los Instrumentos
- Datos con Persistencia

Sesgo de Nickell

- El modelo típico de datos de panel macro incluye valores rezagados de la variable y_{it} entre los regresores. En tales modelos es claro que solo la condición de exogeneidad secuencial puede asumirse. Considere, por ejemplo, $x_{it} = y_{it-1}$:

$E(u_{it} | c_i, y_{i1}, \dots, y_{it-2}, y_{it-1})$ podría ser igual a 0
pero $E(u_{it} y_{it})$ no porque y_{it} depende de u_{it}

- El ejemplo de arriba es conocido como el **modelo AR(1) de efectos no observados**, el modelo de panel dinámico por excelencia:

$$y_{it} = c_i + \rho y_{it-1} + u_{it}$$

Sesgo de Nickell

- Aquí el parámetro de interés es ρ . La expresión analítica del sesgo asintótico de ρ en este modelo fue derivada por Nickell (1981) y se conoce como el **sesgo de Nickell**. Puede escribirse como:

$$\begin{aligned}\nu(\rho, T) &= p \lim \hat{\rho}^{FE} - \rho \\ &= \left\{ \frac{2\rho}{1-\rho^2} - \left[\frac{1+\rho}{T-1} \left(1 - \frac{1}{T} \left(\frac{1-\rho^T}{1-\rho} \right) \right) \right]^{-1} \right\}^{-1}\end{aligned}$$

- Nickell mostró que este sesgo es siempre negativo si $\rho > 0$ y que nunca converge a cero aún si $\rho = 0$. Más aún, el sesgo se vuelve más grande si se adicionan regresores exógenos a la ecuación. Y los estimadores de los coeficientes que acompañan a los regresores exógenos también son inconsistentes.

- No obstante, este problema ha sido considerado como un problema de T fijo. Con una condición de estabilidad, $|\rho| < 1$, se puede mostrar que:

$$\lim_{T \rightarrow \infty} \nu(\rho, T) = 0$$

Sin embargo, el sesgo persiste aún cuando $T \rightarrow \infty$ si ρ se acerca a uno.

1 Modelos Dinámicos

- Introducción a Modelos de Datos de Panel Dinámicos
- El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
- Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
- El Estimador de Arellano-Bond
- El Estimador de Blundell-Bond
- Extensión: Regresores Exógenos
- Contrastes de Validez de los Instrumentos
- Datos con Persistencia

Arellano-Bond y Blundell-Bond

- Para estimar ρ en forma consistente, necesitamos asumir lo siguiente. Sea:

$$y_i^{t-1} = \begin{bmatrix} y_{it-1} \\ \vdots \\ y_{i1} \end{bmatrix}$$

el vector de la historia de las observaciones de y_i hasta el período $t - 1$.

- Además del supuesto típico:

$$|\rho| < 1$$

vamos a asumir (y esto es lo máximo que podemos requerir) exogeneidad secuencial:

$$\forall i \in \mathcal{N}, \forall t \in \mathcal{T} : E(u_{it} | c_i, y_i^{t-1}) = 0$$

Arellano-Bond and Blundell-Bond

- La exogeneidad secuencial implica la ausencia de correlación serial:

$$\begin{aligned}\forall s > 0 : \\ E(u_{it}u_{it-s}) &= E[E(u_{it}u_{it-s}|c_i, y_i^{t-1})] \\ &= E[E(u_{it}(y_{it-s} - c_i + \rho y_{it-s-1})|c_i, y_i^{t-1})] \\ &= E[(y_{it-s} - c_i + \rho y_{it-s-1})E(u_{it}|c_i, y_i^{t-1})] \\ &= E[(y_{it-s} - c_i + \rho y_{it-s-1})0] \\ &= 0\end{aligned}$$

Arellano-Bond and Blundell-Bond

- Volviendo a la condición de estacionariedad, en el contexto de series temporales, usualmente se asume que el proceso estocástico comienza arbitrariamente muy lejos en el tiempo. Esto implica que $|\rho| < 1$ asegura la estacionariedad. Sin embargo, la condición de estabilidad no asegura la estacionariedad si el proceso *comienza* en algún período finito t , por ejemplo $t = 1$. Para ver esto, escribamos el proceso en términos de esta condición inicial:

$$\begin{aligned}y_{it} &= c_i + \rho y_{it-1} + u_{it} \\&= c_i + \rho (c_i + \rho y_{it-2} + u_{it-1}) + u_{it} \\&= (1 + \rho) c_i + \rho^2 y_{it-2} + u_{it} + \rho u_{it-1} \\&\vdots \\y_{it} &= c_i \left(\sum_{s=0}^{t-2} \rho^s \right) + \rho^{t-1} y_{i1} + \sum_{s=0}^{t-2} \rho^s u_{it-s}\end{aligned}$$

Arellano-Bond and Blundell-Bond

- Como no podemos seguir con este proceso recursivo hasta llegar a $-\infty$ para eliminar el término $\rho^{t-1}y_{i1}$, encontramos que la distribución de y_{it} depende de la distribución de la observación inicial, y_{i1} . por lo tanto, los momentos de y_{it} dependen de los momentos de y_{i1} .
- Tomemos primero la esperanza de y_{it} condicional en c_i , que es:

$$\begin{aligned} E(y_{it}|c_i) &= E[E(y_{it}|y_i^{t-1}, c_i)|c_i] = \\ E\left[E\left(c_i\left(\sum_{s=0}^{t-2}\rho^s\right) + \rho^{t-1}y_{i1} + \sum_{s=0}^{t-2}\rho^s u_{it-s}|y_i^{t-1}, c_i\right)|c_i\right] &= \\ c_i\left(\sum_{s=0}^{t-2}\rho^s\right) + \rho^{t-1}E[y_{i1}|c_i] &= \\ c_i\frac{1-\rho^{t-1}}{1-\rho} + \rho^{t-1}E[y_{i1}|c_i] \end{aligned}$$

Arellano-Bond and Blundell-Bond

- Este momento depende de la misma esperanza pero de y_{i1} . Para un proceso AR(1) estacionario, condicional en c_i , la esperanza hubiera sido:

$$\mu_i = \frac{c_i}{1 - \rho}$$

- Para que esta esperanza sea la esperanza de nuestro proceso de panel, necesitamos asumir además de $|\rho| < 1$ que:

$$E[y_{i1}|c_i] = \mu_i = \frac{c_i}{1 - \rho}$$

Arellano-Bond and Blundell-Bond

- Bajo este supuesto adicional, se verifica que μ_i es de hecho la esperanza de y_{it} para cada t :

$$\begin{aligned} E(y_{it}|c_i) &= c_i \frac{1 - \rho^{t-1}}{1 - \rho} + \rho^{t-1} E[y_{i1}|c_i] \\ &= c_i \frac{1 - \rho^{t-1}}{1 - \rho} + \rho^{t-1} \frac{c_i}{1 - \rho} \\ &= \frac{c_i}{1 - \rho} (1 - \rho^{t-1} + \rho^{t-1}) \\ &= \frac{c_i}{1 - \rho} = \mu_i \end{aligned}$$

- Lo mismo ocurre con las autocovarianzas del proceso de panel. Asumamos que:

$$\forall i \in \mathcal{N}, \forall t \in \mathcal{T} : E(u_{it}^2 | c_i, y_i^{t-1}) = \sigma^2$$

Arellano-Bond and Blundell-Bond

- y que:

$$\text{var}(y_{i1}|c_i) = \frac{\sigma^2}{1 - \rho^2}$$

- Ahora, bajo los supuestos hechos hasta ahora, el modelo AR(1) de panel puede escribirse en términos de las desviaciones con respecto a su media como sigue:

$$\begin{aligned} y_{it} - \mu_i &= \rho(y_{it-1} - \mu_i) + u_{it} \\ &= \rho^{t-1}(y_{i1} - \mu_i) + \sum_{s=0}^{t-2} \rho^s u_{it-s} \\ y_{it-k} - \mu_i &= \rho(y_{it-k-1} - \mu_i) + u_{it-k} \\ &= \rho^{t-k-1}(y_{i1} - \mu_i) + \sum_{s=0}^{t-k-2} \rho^s u_{it-k-s} \end{aligned}$$

Arellano-Bond and Blundell-Bond

- Entonces, las autocovarianzas pueden escribirse como:

$$\begin{aligned} cov(y_{it}, y_{it-k} | c_i) &= E[(y_{it} - \mu_i)(y_{it-k} - \mu_i) | c_i] \\ &= \rho^{2t-k-2} var(y_{i1} | c_i) + \rho^k \sum_{s=0}^{t-k-2} \rho^{2s} \sigma^2 \\ &= \rho^{2t-k-2} \frac{\sigma^2}{1 - \rho^2} + \rho^k \frac{1 - \rho^{2t-2k-2}}{1 - \rho^2} \sigma^2 \\ &= \frac{\sigma^2}{1 - \rho^2} (\rho^{2t-k-2} + \rho^k - \rho^{2t-k-2}) \\ &= \rho^k \frac{\sigma^2}{1 - \rho^2} = \gamma_k \end{aligned}$$

Arellano-Bond and Blundell-Bond

- Por lo tanto, para cada t tenemos:

$$\begin{aligned} \text{var}(y_{it}|c_i) &= \gamma_0 \\ &= \frac{\sigma^2}{1 - \rho^2} \\ &= \text{var}(y_{i1}|c_i) \end{aligned}$$

- Para concluir, si los momentos de la primera observación del proceso son los momentos en estado estacionario, el proceso entero es estacionario y comparte esos mismos momentos.

1 Modelos Dinámicos

- Introducción a Modelos de Datos de Panel Dinámicos
- El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
- Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
- El Estimador de Arellano-Bond
- El Estimador de Blundell-Bond
- Extensión: Regresores Exógenos
- Contrastes de Validez de los Instrumentos
- Datos con Persistencia

Estimador de Arellano-Bond

- Volviendo al problema de la estimación, las transformaciones usuales para eliminar c_i generan estimadores inconsistentes en el contexto de paneles dinámicos. El problema recae en el hecho de que, bajo exogeneidad secuencial, tomar primeras diferencias (o transformar por efectos fijos) elimina c_i pero provoca que la condición de exogeneidad se viole:

$$\begin{aligned}y_{it} &= c_i + \rho y_{it-1} + u_{it} \\ \Delta y_{it} &= \rho \Delta y_{it-1} + \Delta u_{it} \\ E(\Delta y_{it-1} \Delta u_{it}) &= E(y_{it-1} u_{it} - y_{it-1} u_{it-1} - y_{it-2} u_{it} - y_{it-2} u_{it-1}) \\ &= -E(y_{it-1} u_{it-1}) \neq 0\end{aligned}$$

- Entonces, para estimar ρ en forma consistente, necesitamos instrumentos para Δy_{it-1} . Tomemos y_{it-2} en niveles como instrumento de Δy_{it-1} :

$$E(y_{it-2} \Delta u_{it}) = E(y_{it-2} u_{it} - y_{it-2} u_{it-1})$$

Estimador de Arellano-Bond

- En general,

$$E(y_i^{t-2} \Delta u_{it}) = 0_{t-2}$$

- Esto significa que el instrumento es válido. El estimador de Arellano-Bond es el estimador de GMM con matriz de instrumentos:

$$Z = \begin{bmatrix} Z_1 \\ \vdots \\ Z_N \end{bmatrix}$$
$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \vdots & \vdots & \vdots & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \cdots & y_{iT-2} & \cdots & y_{iT-2} \end{bmatrix}$$

Estimador de Arellano-Bond

- Tomando primeras diferencias e instrumentando con valores rezagados, las primeras dos observaciones no pueden usarse en la estimación. Definamos los vectores de observaciones **usables** como:

$$\Delta y = \begin{bmatrix} \Delta y_{13} \\ \vdots \\ \Delta y_{NT} \end{bmatrix} ; \quad \Delta y_{-1} = \begin{bmatrix} \Delta y_{12} \\ \vdots \\ \Delta y_{N(T-1)} \end{bmatrix}$$

y matriz ponderadora W .

- Entonces, el estimador de Arellano-Bond de ρ es:

$$\widehat{\rho}^{AB}(W) = \left[\Delta' y_{-1} (ZWZ') \Delta y_{-1} \right]^{-1} \Delta' y_{-1} (ZWZ') \Delta y$$

Estimador de Arellano-Bond

- Hay dos elecciones disponibles para W : W_1 , una matriz ponderadora de un solo paso, y W_2 , una matriz ponderadora de dos pasos que usa los residuos del estimador que usa W_1 , $\hat{u}_i^{(1)}$:

$$W_1 = \left(\sum_{i=1}^N \Delta Z_i' \Delta Z_i \right)^{-1}$$
$$W_2 = \left(\sum_{i=1}^N Z_i' \Delta \hat{u}_i^{(1)} \Delta \hat{u}_i^{(1)'} Z_i \right)^{-1}$$

Estimador de Arellano-Bond

- Cada matriz ponderadora da un estimador de Arellano-Bond diferente:

$$\hat{\rho}_1^{AB} = \left[\Delta' y_{-1} (ZW_1 Z') \quad \Delta y_{-1} \right]^{-1} \Delta' y_{-1} (ZW_1 Z') \quad \Delta y$$

$$\hat{\rho}_2^{AB} = \left[\Delta' y_{-1} (ZW_2 Z') \quad \Delta y_{-1} \right]^{-1} \Delta' y_{-1} (ZW_2 Z') \quad \Delta y$$

- La estimación de las varianzas de estos estimadores es:

$$\widehat{var}(\hat{\rho}_1^{AB}) = \hat{\sigma}^2 \left[\Delta' y_{-1} (ZW_1 Z') \quad \Delta y_{-1} \right]^{-1}$$

$$\widehat{var}(\hat{\rho}_2^{AB}) = \hat{\sigma}^2 \left[\Delta' y_{-1} (ZW_2 Z') \quad \Delta y_{-1} \right]^{-1}$$

donde $\hat{\sigma}^2$ es la estimación de la varianza de u_{it} .

Estimador de Arellano-Bond

- Una estimación consistente de σ^2 viene dada por:

$$\hat{\sigma}^2 = \frac{1}{2(N(T-2) - K)} \sum_{i=1}^N \widehat{\Delta u_i}' \widehat{\Delta u_i}$$

donde K es la dimensión de $X_i \equiv y_{it-1} = 1$ y $\widehat{\Delta u_i}$ es la estimación de los errores del modelo transformado. Note que el número de observaciones en la ecuación estimada es $N(T-2)$ porque perdemos las primeras dos observaciones debido a la variable dependiente rezagada y a la transformación de diferencias finitas.

- $$\widehat{\Delta u_i} = \Delta y_i - \Delta y_{it-1} \hat{\rho}_1^{AB}$$

1 Modelos Dinámicos

- Introducción a Modelos de Datos de Panel Dinámicos
- El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
- Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
- El Estimador de Arellano-Bond
- **El Estimador de Blundell-Bond**
- Extensión: Regresores Exógenos
- Contrastes de Validez de los Instrumentos
- Datos con Persistencia

Estimador de Blundell-Bond

- Un buen estimador de variables instrumentales es aquel en el que el instrumento y la variable instrumentada están altamente correlacionadas. Esta correlación, en el caso del estimador de Arellano-Bond, está dada por:

$$\begin{aligned} E(y_{it-2}\Delta y_{it-1}) &= E[y_{it-2}(c_i + (\rho - 1)y_{it-2} + u_{it})] \\ &= E[y_{it-2}c_i] + (\rho - 1)E[y_{it-2}^2] \end{aligned}$$

- Como asumiremos más adelante, $E[y_{it-2}c_i] = 0$. Entonces,

$$E(y_{it-2}\Delta y_{it-1}) = (\rho - 1)E[y_{it-2}^2]$$

- Cuando el proceso de panel es un modelo de panel con raíz unitaria, esto es, cuando $\rho = 1$, esta correlación alcanza su menor valor (en valor absoluto). Lo que sucede es que, bajo $\rho = 1$, el proceso es un paseo aleatorio (random walk) de panel y sus incrementos, Δy_{it} son ruido blanco.

Estimador de Blundell-Bond

- Entonces, cuando ρ se aproxima a uno, los instrumentos utilizados por Arellano-Bond se vuelven instrumentos débiles.
- Por lo tanto, Blundell-Bond proponen agregar nuevos instrumentos.
- Para esto, note que los valores rezagados de la diferencia Δy_{it} son ortogonales a los niveles de u_{it} :

$$\forall i \in \mathcal{N}, \forall s > 0 : E(\Delta y_{it-s} u_{it}) = 0$$

- Entonces, podemos estimar el siguiente modelo:

$$\begin{cases} \Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it} \\ y_{it} = \rho y_{it-1} + v_{it} \end{cases}$$

$$\text{con } v_{it} = c_i + u_{it}.$$

Estimador de Blundell-Bond

- Usando los instrumentos de Arellano-Bond para la ecuación en diferencias y las diferencias rezagadas como instrumento para la ecuación en niveles.
- Para obtener estimadores consistentes, el nuevo conjunto de instrumentos no tiene que estar correlacionado con el nuevo término de error, v_{it} . Esto está asegurado por el supuesto:

$$\forall i \in \mathcal{N}, \forall s > 0 : E(\Delta y_{it-s} c_i) = 0$$

- Note que esto implica la condición, ya establecida, que:

$$E[y_{it-2} c_i] = 0$$

- Para obtener una única expresión para el modelo y el estimador, note que $\Delta u_{it} = \Delta v_{it}$.

Estimador de Blundell-Bond

- Además, definamos D como la matriz $(T - 2) \times (T - 1)$ que representa la operación de tomar primeras diferencias. También, definamos Z^+ como la matriz de instrumentos aumentada:

$$Z_i^+ = \begin{bmatrix} Z_i & 0 & 0 & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & 0 & 0 & \dots & 0 \\ 0 & 0 & \Delta y_{i3} & 0 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & 0 \\ 0 & 0 & 0 & \dots & \ddots & 0 \\ 0 & 0 & 0 & \dots & 0 & \Delta y_{i(T-1)} \end{bmatrix}$$
$$Z^+ = \begin{bmatrix} Z_1^+ \\ \vdots \\ Z_N^+ \end{bmatrix}$$

Estimador de Blundell-Bond

- y H como la matriz, en bloques, de transformación:

$$H = \begin{bmatrix} D \\ I_{T-1} \end{bmatrix}$$

- El modelo en

$$\begin{cases} \Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it} \\ y_{it} = \rho y_{it-1} + v_{it} \end{cases}$$

- puede ser escrito como:

$$(I_N \otimes H) y = \rho (I_N \otimes H) y_{-1} + (I_N \otimes H) v$$

Estimador de Blundell-Bond

- con:

$$\begin{aligned} y_i &= \begin{bmatrix} y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} ; \quad y_{i-1} = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{i(T-1)} \end{bmatrix} ; \quad v_i = \begin{bmatrix} v_{i2} \\ \vdots \\ v_{iT} \end{bmatrix} \\ y &= \begin{bmatrix} y_2 \\ \vdots \\ y_N \end{bmatrix} ; \quad y_{-1} = \begin{bmatrix} y_{11} \\ \vdots \\ y_{N(T-1)} \end{bmatrix} ; \quad v = \begin{bmatrix} v_2 \\ \vdots \\ v_N \end{bmatrix} \end{aligned}$$

- Entonces, queremos estimar ρ por GMM con la matriz de instrumentos aumentada. Nuevamente, necesitamos la condición de ortogonalidad:

$$E \left[Z^{+'} (I_N \otimes H) v \right] = 0_{2N(T-2)}$$

Estimador de Blundell-Bond

- El estimador de **Blundell-Bond** es el siguiente estimador de dos pasos.
- En el primer paso calculamos el siguiente estimador de GMM:

$$\begin{aligned}\hat{\rho}^{(1)} &= \left\{ \left[y'_{-1} \left(I_N \otimes H' \right) Z^+ \right]^{-1} W^{(1)} \left[Z^{+'} \left(I_N \otimes H \right) y_{-1} \right] \right\}^{-1} \\ &\times \left\{ \left[y'_{-1} \left(I_N \otimes H' \right) Z^+ \right]^{-1} W^{(1)} \left[Z^{+'} \left(I_N \otimes H \right) y \right] \right\}\end{aligned}$$

- con:

$$W^{(1)} = \left[\sum_{i=1}^N Z_i^{+'} H H' Z_i^+ \right]^{-1}$$

Estimador de Blundell-Bond

- Este estimador $\hat{\rho}^{(1)}$ se conoce como el **estimador de Blundell-Bond de un paso**.
- La varianza del estimador de Blundell-Bond de un paso es:

$$\widehat{var} \left(\hat{\rho}^{(1)} \right) = \hat{\sigma}^2 \left\{ \left[y_{-1}' \left(I_N \otimes H' \right) Z^+ \right]^{-1} W^{(1)} \left[Z^{+'} \left(I_N \otimes H \right) y_{-1} \right] \right\}^{-1}$$

- Una estimación consistente de σ^2 viene dada por:

$$\hat{\sigma}^2 = \frac{1}{2(N(T-2) - K)} \sum_{i=1}^N \widehat{\Delta u_i}' \widehat{\Delta u_i}$$

- Con

$$\widehat{\Delta u_i} = \Delta y_i - \Delta y_{it-1} \hat{\rho}^{(1)}$$

Estimador de Blundell-Bond

- Tome los residuos de este primer paso, $\widehat{v}_i^{(1)}$. El estimador propuesto por Blundell-Bond es el estimador de GMM con una matriz ponderadora óptima que puede obtenerse en el segundo paso:

$$\widehat{\rho}^{BB} = \left\{ \left[y'_{-1} (I_N \otimes H') Z^+ \right]^{-1} W_{BB} \left[Z^{+'} (I_N \otimes H) y_{-1} \right] \right\}^{-1} \\ \times \left\{ \left[y'_{-1} (I_N \otimes H') Z^+ \right]^{-1} W_{BB} \left[Z^{+'} (I_N \otimes H) y \right] \right\}$$

- con:

$$W_{BB} = \left[\sum_{i=1}^N Z_i^{+'} H \widehat{v}_i \widehat{v}_i' H' Z_i^+ \right]^{-1}$$

Estimador de Blundell-Bond

- La varianza del estimador de Blundell-Bond es:

$$\widehat{var}(\widehat{\rho}^{BB}) = \hat{\sigma}^2 \left\{ \left[y_{-1}' (I_N \otimes H') Z^+ \right]^{-1} W_{BB} \left[Z^{+'} (I_N \otimes H) y_{-1} \right] \right\}^{-1}$$

- Una estimación consistente de σ^2 viene dada por:

$$\hat{\sigma}^2 = \frac{1}{2(N(T-2) - K)} \sum_{i=1}^N \widehat{\Delta u}_i' \widehat{\Delta u}_i$$

donde K es la dimensión de $X_i \equiv y_{it-1} = 1$ y $\widehat{\Delta u}_i$ es la estimación de los errores del modelo transformado. Note que el número de observaciones en la ecuación estimada es $N(T-2)$ porque perdemos las primeras dos observaciones debido a la variable dependiente rezagada y a la transformación de diferencias finitas.

$$\widehat{\Delta u}_i = \Delta y_i - \Delta y_{it-1} \widehat{\rho}^{BB}$$

1 Modelos Dinámicos

- Introducción a Modelos de Datos de Panel Dinámicos
- El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
- Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
- El Estimador de Arellano-Bond
- El Estimador de Blundell-Bond
- Extensión: Regresores Exógenos
- Contrastes de Validez de los Instrumentos
- Datos con Persistencia

AB y BB con Regresores Exógenos

- Los estimadores de Arellano-Bond y de Blundell-Bond pueden extenderse en forma directa a modelos que incluyan regresores **estrictamente exógenos**, agrupados en el vector k -dimensional x_{it} . tales modelos se denominan **AR(1)-X modelos de efectos no observados**:

$$y_{it} = c_i + \rho y_{it-1} + \gamma' x_{it} + u_{it}$$

where:

$$\forall i \in \mathcal{N}, \forall t \in \mathcal{T} : E(u_{it} | c_i, x_{i1}, x_{i2}, \dots, x_{iT}) = 0$$

- La extensión es directa ya que los regresores estrictamente exógenos, x_{it} pueden ser usados como sus propios instrumentos. Por lo tanto, aumentando adecuadamente las matrices Z, Z^+ , las mismas fórmulas de arriba pueden ser usadas para estimar consistentemente ρ y γ .

AB y BB con Regresores Exógenos

- Si los regresores x_{it} son secuencialmente exógenos, esto es $E(x_{it} u_{is}) = 0, \forall t \neq s$, entonces solo $(x_{i1}, \dots, x_{is-1})$ son instrumentos válidos para la ecuación en primeras diferencias del período s .
- Entonces, el estimador de Arellano-Bond utiliza como matriz de instrumentos:

$$Z_i = \begin{bmatrix} y_{i1} & x_{i2} & 0 & 0 & \dots & 0 \\ 0 & y_{i1} & y_{i2} & x_{i2} & x_{i3} & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & y_{iT-2} & x_{i2} & \dots & x_{iT-1} \end{bmatrix}$$

- En la ecuación $\Delta y_{it} = \Delta x_{it}^\dagger \delta + \Delta u_{it}, \quad t = 3, 4, \dots, T$. Con $x_{it}^\dagger = [y_{it-1} \ x_{it}]$.

AB y BB con Regresores Exógenos

- Claramente, las x_{it} pueden tener elementos estrictamente exógenos o secuencialmente exógenos en cuyo caso la matriz de instrumentos se puede definir de forma apropiada.
- Sin embargo, el número de columnas de Z_i en cualquiera de los casos anteriores puede llegar a ser muy grande produciéndose una pérdida de eficiencia.
- Para corregir esto, en general, los programas que resuelven estos modelos utilizan solo algunas de esas columnas.

AB y BB con Regresores Exógenos: ejemplo

- Considere el siguiente modelo,

$$y_{it} = \rho y_{it-1} + x_{it}\beta + c_i + u_{it}, \quad T = 5.$$

- Donde x_{it} es 1×1 , y es una variable estrictamente exógena.
- Primero aplicamos diferencias finitas.

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta x_{it}\beta + \Delta u_{it},$$

- Ahora instrumentemos la ecuación utilizando solo dos rezagos. Entonces Z_i queda,

$$Z_i = \begin{bmatrix} y_{i1} & 0 & 0 & 0 & 0 & \Delta x_{i3} \\ 0 & y_{i1} & y_{i2} & 0 & 0 & \Delta x_{i4} \\ 0 & 0 & 0 & y_{i2} & y_{i3} & \Delta x_{i5} \end{bmatrix}$$

- En este caso como las x 's son estrictamente exógenas utilizamos la propia variable como instrumento de si misma.

AB y BB con Regresores Exógenos: ejemplo

- Note que debido al rezago de la variable dependiente se pierde una observación y debido a las diferencias finitas se pierde otra observación de forma que $t = 3, 4, 5$.
- Escribiendo el modelo *stacking* sobre t , tenemos:

$$Dy_i = Dy_i^{(-1)}\rho + DX_i\beta + Du_i$$

- Reagrupando:

$$Dy_i = [Dy_i^{(-1)} | DX_i](\rho \ \beta)' + Du_i$$

- Llamando $V_i = D[y_i^{(-1)} | X_i]$, obtenemos

$$\begin{pmatrix} \hat{\rho} \\ \hat{\beta} \end{pmatrix} = \left[\left(\sum_{i=1}^N V_i' DZ_i \right) W \left(\sum_{i=1}^N Z_i' D V_i \right) \right]^{-1} \\ \times \left(\sum_{i=1}^N V_i' DZ_i \right) W \left(\sum_{i=1}^N Z_i' D y_i \right)$$

AB y BB con Regresores Exógenos: ejemplo

- Igual que antes, el estimador de un paso, usa

$$W = \left(\frac{1}{N} \sum_{i=1}^N (DZ_i)' DZ_i \right)^{-1}$$

- Si queremos estimar usando el estimador de Blundell-Bond, entonces hay que agregar a las ecuaciones en primeras diferencias, las ecuaciones en niveles,

$$\begin{aligned} Dy_i &= Dy_i^{(-1)}\rho + DX_i\beta + Du_i \\ y_i &= y_i^{(-1)}\rho + X_i\beta + c_iJ_T + u_i \end{aligned}$$

- Y estimar usando la matriz de instrumentos,

$$Z_i^+ = \begin{bmatrix} Z_i & 0 & 0 & 0 & 0 & 0 \\ 0 & \Delta y_{i2} & 0 & 0 & x_{i3} & 1 \\ 0 & 0 & \Delta y_{i3} & 0 & x_{i4} & 1 \\ 0 & 0 & 0 & \Delta y_{i4} & x_{i5} & 1 \end{bmatrix}$$

1 Modelos Dinámicos

- Introducción a Modelos de Datos de Panel Dinámicos
- El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
- Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
- El Estimador de Arellano-Bond
- El Estimador de Blundell-Bond
- Extensión: Regresores Exógenos
- Contrastes de Validez de los Instrumentos
- Datos con Persistencia

Contrastes de Validez de los Instrumentos

- Como ambos estimadores, el de Arellano-Bond y el de Blundell-Bond son estimadores de GMM, es usual contrastar por la validez de sus instrumentos. Para el contexto de datos de panel hay dos contrastes disponibles.
- El primer test es el test- J de Sargan típico. Este es el test estándar de validez de los instrumentos. Tome una matriz de instrumentos J y los errores en primeras diferencias, Δu_i . Las hipótesis son:

$$\begin{cases} H_0 : E(J_i \Delta u_i) = 0_{T-2} \\ H_1 : E(J_i \Delta u_i) \neq 0_{T-2} \end{cases}$$

- El estadístico de contraste, $s(J)$, es:

$$s(J) = \left(\sum_{i=1}^N \Delta \hat{u}_i^{(J)'} J_i \right) \left(\sum_{i=1}^N J_i' \Delta \hat{u}_i^{(J)} \Delta \hat{u}_i^{(J)'} J_i \right)^{-1} \left(\sum_{i=1}^N J_i' \Delta \hat{u}_i^{(J)} \right)$$

Contrastes de Validez de los Instrumentos

- Se puede mostrar que, para $J = Z$, $J = Z^+$:

$$s^{AB} = s(Z) \xrightarrow{\mathcal{D}} \chi^2_{\text{col}(Z)-k}$$

$$s^{BB} = s(Z^+) \xrightarrow{\mathcal{D}} \chi^2_{\text{col}(Z^+)-k}$$

with $\text{col}(Z^+) = T - 2 + \text{col}(Z)$.

- Rechazar la hipótesis nula en el test- J significa que los instrumentos no son válidos. Esto implica que el DGP no es el modelo AR(1) de efectos no observables ya que para este modelo los instrumentos de Arellano-Bond y Blundell-Bond son válidos. Entonces, el test- J puede ser pensado como un test de **especificación del modelo**.

Contrastes de Validez de los Instrumentos

- El segundo test es el test-M. A diferencia del test anterior, este contraste es un test de especificación **directamente**.
- Tome el modelo en primeras diferencias. Si los errores en niveles son ruido blanco, los errores en primeras diferencias tendrán una estructura de autocovarianzas determinada.
- Más precisamente, la autocovarianza de primer orden de los errores en primeras diferencias es negativa y la autocovarianza de segundo orden es cero:

$$\begin{aligned}\gamma_{\Delta u}(1) &= E(\Delta u_{it} \Delta u_{it-1}) \\ &= E[(u_{it} - u_{it-1})(u_{it-1} - u_{it-2})] \\ &= E[u_{it}u_{it-1} - u_{it-1}u_{it-1} - u_{it}u_{it-2} + u_{it-1}u_{it-2}] \\ &= -\gamma_u(0) < 0 \\ \gamma_{\Delta u}(2) &= E(\Delta u_{it} \Delta u_{it-2}) \\ &= E[(u_{it} - u_{it-1})(u_{it-2} - u_{it-3})] \\ &= 0\end{aligned}$$

Contrastes de Validez de los Instrumentos

- Entonces, tenemos los siguientes dos conjuntos de hipótesis:

$$(1) \begin{cases} H_0^{(1)} : \gamma_{\Delta u}(1) = 0 \\ H_1^{(1)} : \gamma_{\Delta u}(1) < 0 \end{cases}$$

$$(2) \begin{cases} H_0^{(2)} : \gamma_{\Delta u}(2) = 0 \\ H_1^{(2)} : \gamma_{\Delta u}(2) \neq 0 \end{cases}$$

y debemos rechazar $H_0^{(1)}$ y aceptar $H_0^{(2)}$. Los estadísticos de contraste, m_1 y m_2 , son asintóticamente normales bajo cada hipótesis nula (y son provistos por Stata después de la estimación):

$$m_1 \xrightarrow{\mathcal{D}H_0^{(1)}} \mathcal{N}(0, 1)$$

$$m_2 \xrightarrow{\mathcal{D}H_0^{(2)}} \mathcal{N}(0, 1)$$

1 Modelos Dinámicos

- Introducción a Modelos de Datos de Panel Dinámicos
- El Modelo AR(1) de Efectos No Observados y el sesgo de Nickell
- Estimación de ρ Consistente: Arellano-Bond y Blundell-Bond
- El Estimador de Arellano-Bond
- El Estimador de Blundell-Bond
- Extensión: Regresores Exógenos
- Contrastes de Validez de los Instrumentos
- Datos con Persistencia

Estimación LSDV Corregida (Kiviet)

- Considere el siguiente modelo de componentes no observados,

$$y_{it} = \gamma y_{it-1} + x_{it}\beta + c_i + u_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T \quad (3)$$

con,

- ▶ $c_i \sim N(0, \sigma_c^2)$, $\sigma_c^2 > 0$; $u_{it} \sim N(0, \sigma_u^2)$, $\sigma_u^2 > 0$
- ▶ (i) $E(u_{it}u_{js}) = 0$ $j \neq i$ o $t \neq s$; (ii) $E(c_i c_j) = 0$ $j \neq i$
- ▶ (iii) $E(c_i u_{jt}) = 0 \quad \forall j, i, t$; (iv) $E(x'_{it} u_{js}) = 0 \quad \forall j, i, t, s$
- ▶ (v) $E(x'_{jt} c_i) = \text{desconocida} \quad \forall j, i, t$
- ▶ y_{i0} es una variable aleatoria con (vi) $E(y_{i0} u_{jt}) = 0 \quad \forall j, i, t$; (vii) $E(y_{i0} c_j) = \text{desconocida} \quad \forall j, i$.
- ▶ (viii) $E(w_{it} u_{jt}) = 0, \quad \forall j, i, t$, donde $w_{it} = y_{it} - \frac{1}{1-\gamma} c_i$.
- Kiviet muestra que el sesgo del estimador de FE puede aproximarse con un error de tamaño $O_p(N^{-1}T^{-3/2})$.
- Asuma que los supuestos (i) a (viii) y $|\gamma| < 1$ se cumplen. Entonces,

Estimación LSDV Corregida (Kiviet)

- Teorema 1 (Kiviet, 1995 pp 64)

$$\begin{aligned} E(\hat{\delta}_{FE} - \delta) &= -\sigma_u^2 \bar{D}^{-1} \left(\frac{N}{T} (J_T' C J_T) [2q - \bar{W}' Q_{NT} \bar{W} \bar{D}^{-1} q] \right. \\ &\quad + \text{tr}\{\bar{W}' (I_N \otimes Q_T C Q_T) \bar{W} \bar{D}^{-1}\} q \\ &\quad + \bar{W}' (I_N \otimes Q_T C Q_T) \bar{W} \bar{D}^{-1} q + \sigma_u^2 N q' \bar{W}' \bar{D}^{-1} q \\ &\quad \times \left[\frac{N}{T} (J_T' C J_T) \text{tr}\{C' Q_T C\} + 2 \text{tr}\{C' Q_T C Q_T C\} \right] \Bigg) \\ &\quad + O_p(N^{-1} T^{-3/2}) \\ &= \text{Sesgo}_{\text{Kiviet}} + O_p(N^{-1} T^{-3/2}) \end{aligned}$$

donde $\bar{D} = \bar{W}' Q_{NT} \bar{W} + \sigma_u^2 N \text{tr}\{C' Q_T C\} q q'$; $Q_{NT} \bar{W} = E(Q_{NT} W)$;
 $q = (1 \ 0 \ \dots \ 0)'$; $\delta' = (\gamma \ \beta')$ y,

Estimación LSDV Corregida (Kiviet)

$$C = \begin{bmatrix} 0 & 0 & \dots & \dots & \dots & \dots & 0 \\ 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ \gamma & 1 & 0 & \dots & \dots & \dots & 0 \\ \gamma^2 & \gamma & 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma^{T-2} & \dots & \dots & \dots & \gamma & 1 & 0 \end{bmatrix}$$

- El único parámetro desconocido en C es γ .
- Kiviet sugiere reemplazarlo con la estimación de IV de Anderson-Hsiao.
- Para el modelo $AR(1)$ Anderson y Hsiao (1982) aplican diferencias finitas para eliminar c_i .
- $\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}, \quad t > 2.$

Estimación LSDV Corregida (Kiviet)

- Luego, usan POLS IV con instrumentos dados por y_{it-2} o Δy_{it-2} .
- Note que en el período t , todos los elementos de $(y_{it-2}, y_{it-3}, \dots, y_{i0})$ son instrumentos válidos porque Δu_{it} , no está correlacionada con y_{it-h} , $h \geq 2$.
- Como el estimador de Anderson y Hsiao no usa todos los instrumentos disponibles no es completamente eficiente.
- El estimador propuesto por Kiviet es entonces:
 1. Estimar el modelo por el método de LSDV y obtener la estimación no consistente de δ , $\hat{\delta}_{FE}$.
 2. Calcular la estimación consistente como: $\hat{\delta}_{LSDVC} = \hat{\delta}_{FE} - \text{Sesgo}_{\text{Kiviet}}$.

Estimación LSDV Corregida (Kiviet)

- Para estimar la matriz de varianzas y covarianzas de los coeficientes de Kiviet siga los siguientes pasos:

1. Calcular $\hat{u}_i^{LSDVC} = \ddot{y}_i - \ddot{Z}_i \hat{\delta}_{LSDVC}$ donde \ddot{Z}_i incluye \ddot{y}_{it-1} y $\ddot{x}_{i,t}$.
2. Estime

$$\hat{\sigma}_u^2 = \frac{1}{NT - N - T - K + 1} \sum_{i=1}^N \hat{u}_i^{LSDVC'} \hat{u}_i^{LSDVC}$$

3. $Var(\hat{\delta}_{LSDVC}) = \hat{\sigma}_u^2 \left(\sum_{i=1}^N \ddot{Z}_i' \ddot{Z}_i \right)^{-1}$

Econometría de Datos de Panel

Maestrías en Economía y Econometría

Lecture 3

- 1 Sesgo de Selección: “Attrition” y Truncamiento Incidental
 - Cuándo Podemos Ignorar el Sesgo de Selección?
 - Contrastes por Sesgo de Selección
 - Corrección del Sesgo de Selección Muestral: Truncamiento Incidental
 - Procedimiento de Wooldridge
 - Procedimiento de Rochina-Barrachina
 - Corrección del Sesgo de Selección Muestral: Attrition

- 2 Sesgo de Selección y Endogeneidad

- 1 Sesgo de Selección: “Attrition” y Truncamiento Incidental
 - Cuándo Podemos Ignorar el Sesgo de Selección?
 - Contrastes por Sesgo de Selección
 - Corrección del Sesgo de Selección Muestral: Truncamiento Incidental
 - Procedimiento de Wooldridge
 - Procedimiento de Rochina-Barrachina
 - Corrección del Sesgo de Selección Muestral: Attrition
- 2 Sesgo de Selección y Endogeneidad

Paneles No Balanceados

- Muchas veces, los datos que tenemos tienen la característica de que algunas observaciones de series temporales no están disponibles para algunas observaciones de corte transversal.
- Cuando esto ocurre, decimos que tenemos **paneles no balanceados**.
- Los paneles no balanceados pueden surgir por varias razones.
- Primero, por diseño de la muestra. Por ejemplo, el procedimiento puede simplemente rotar algunas de las observaciones de corte transversal de acuerdo a una regla específica (**paneles rotativos**).
- Un problema más complicado surge cuando algunas unidades de corte transversal eligen salirse del panel (**attrition**).
- Un problema diferente es cuando las unidades no desaparecen del panel pero ciertas variables no son observadas por al menos algunos períodos temporales (**truncamiento incidental**).

Paneles No Balanceados

- Cualquiera de estos casos puede presentar potencialmente un problema de sesgo de selección muestral.
- Si la decisión de rotar las unidades de corte transversal no se hace aleatoriamente, ó si hay no respuesta relacionada con la variable a explicar, tendremos un problema de sesgo en la muestra.
- De la misma manera, si la attrition se basa en factores sistemáticamente relacionados con la variable a explicar, entonces tendremos un problema de sesgo de selección.
- Por último, si la variable a explicar solo se observa para algunos valores determinados por el comportamiento de alguna otra variable, entonces, potencialmente habrá sesgo de selección.

Paneles No Balanceados

- Comenzaremos analizando los supuestos bajo los cuales, el estimador usual de FE es consistente en paneles desbalanceados.
- Considere el siguiente modelo de componentes no observados,

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1, 2, \dots, T \quad (1)$$

donde x_{it} es $1 \times K$ y β es $K \times 1$, y $\text{Cov}(x_{it}, c_i) \neq 0$. Asumimos que hay disponibles N observaciones de corte transversal y que la teoría asintótica relevante es con $N \rightarrow \infty$.

- Considere el caso en el que algunos períodos temporales no se encuentran disponibles para algunas unidades de corte transversal. Piense en $t = 1$ como el primer período temporal para el que existen datos para toda la población y en $t = T$ como el último período temporal.

Paneles No Balanceados

- Para una selección al azar i desde la población, sea $s_i \equiv (s_{i1}, s_{i2}, \dots, s_{iT})'$ el vector $T \times 1$ de indicadores de selección: $s_{it} = 1$ si $(x_{it}; y_{it})$ es observado, y cero en otro caso.
- Podemos tratar a $(x_i; y_i; s_i) : i = 1; 2; \dots; N$ como una muestra aleatoria de la población; los indicadores de selección nos dicen qué períodos temporales se observan para cada i .
- Podemos encontrar fácilmente supuestos bajo los cuales el estimador de efectos fijos en el panel no balanceado es consistente. Para eso escribamos el estimador como,

$$\begin{aligned}\hat{\beta} &= \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{x}'_{it} \ddot{x}_{it} \right)^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{x}'_{it} \ddot{y}_{it} \right) \\ &= \beta + \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{x}'_{it} \ddot{x}_{it} \right)^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{x}'_{it} u_{it} \right) \quad (2)\end{aligned}$$

Paneles No Balanceados

- Donde $\ddot{x}_{it} = x_{it} - (1/T_i) \sum_{r=1}^T s_{ir} x_{ir}$; $\ddot{y}_{it} = y_{it} - (1/T_i) \sum_{r=1}^T s_{ir} y_{ir}$ y $T_i = \sum_{r=1}^T s_{ir}$.
- Esto es, T_i es el número de períodos temporales observados para la unidad de corte transversal i , y aplicamos la transformación de FE sobre los períodos temporales disponibles.
- Como se desprende de la ecuación anterior, **el estimador de efectos fijos en paneles desbalanceados será consistente siempre que:** $E(s_{it} \ddot{x}'_{it} u_{it}) = 0, \forall t$.
- Como \ddot{x}_{it} depende de todos los elementos en x_i y s_i , necesitamos alguna forma de exogeneidad estricta.
- **Supuesto FEUP.1:** (a) $E(u_{it}|x_i; s_i; c_i) = 0, t = 1; 2; \dots; T$; (b) $\sum_{t=1}^T E(s_{it} \ddot{x}'_{it} \ddot{x}_{it})$ no es singular; y (c) $E(u_i u_i' | x_i; s_i; c_i) = \sigma_u^2 I_T$.
- Bajo el Supuesto FEUP.1 (a), $E(s_{it} \ddot{x}'_{it} u_{it}) = 0$ usando la ley de expectativas iteradas.

Paneles No Balanceados

- FEUP.1 (b) es la condición de rango usual para identificar el estimador de FE después de tomar en cuenta el indicador de selección.
- Estos primeros dos supuestos aseguran la consistencia del estimador de efectos fijos en paneles desbalanceados.
- En el caso de paneles rotativos aleatorios, ó en el caso de que la no respuesta sea aleatoria y en cualquier otro caso en el que el mecanismo de selección sea completamente aleatorio, s_i es independiente de $(u_i; x_i; c_i)$, en cuyo caso el **Supuesto FEUP.1 (a)** se cumple bajo el supuesto estándar de efectos fijos en paneles completos $E(u_{it}|x_i; c_i) = 0 \quad \forall t$.
- En este caso, los supuestos naturales sobre el modelo poblacional implican **consistencia y normalidad asintótica en paneles no balanceados**.

Paneles No Balanceados

- Note que FEUP.1(a) no asume nada acerca de la relación entre s_i y (x_i, c_i) . Por lo tanto, si pensamos que el mecanismo de selección en todos los períodos temporales está correlacionado con c_i ó x_i , pero u_{it} es independiente en media de s_i , dado (x_i, c_i) para todo t , entonces FE en el panel no balanceado es consistente y asintóticamente normal.
- Lo que **FEUP.1(a) descarta es correlación entre s_i , y u_{it}** .
- Si adicionamos el supuesto FEUP.1(c), los procedimientos de inferencia estándar de FE son válidos. En particular bajo FEUP.1 (a) y FEUP.1 (c),

$$\text{Var} \left(\sum_{t=1}^T s_{it} \ddot{x}'_{it} u_{it} \right) = \sigma_u^2 \left[\sum_{t=1}^T E(s_{it} \ddot{x}'_{it} \ddot{x}_{it}) \right]$$

Paneles No Balanceados

- Por lo tanto, la varianza asintótica del estimador de efectos fijos se puede estimar con:

$$\hat{\sigma}_u^2 \left(\sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{X}_{it}' \ddot{X}_{it} \right)^{-1}$$

- El estimador $\hat{\sigma}_u^2$ se puede obtener con,

$$\hat{\sigma}_u^2 = \left[\sum_{t=1}^T (T_i - 1) \right]^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \hat{u}_{it}^2$$

donde \hat{u}_{it} son los residuos de la estimación por efectos fijos.

- Los programas que resuelven datos de panel no balanceados (Stata, EViews etc.) corrigen por grados de libertad restando K de $\sum_{t=1}^T (T_i - 1)$. Se sigue que todos los estadísticos usuales de efectos fijos calculados en un panel no balanceado son válidos.

Paneles No Balanceados

- Rejazar el Supuesto FEUP.1 (c) es fácil: solo se aplica la estimación robusta de la matriz de varianzas y covarianzas al panel desbalanceado.
- Estos resultados implican que el sesgo de selección muestral en el contexto del modelo de FE es un problema sólo si el mecanismo de selección está relacionado con los errores idiosincráticos, u_{it} .
- Por lo tanto, cualquier test de selección muestral solo tiene que contrastar este supuesto.
- La consistencia del estimador de efectos aleatorios (en paneles balanceados o no balanceados) requiere un supuesto adicional, $E(c_i|x_i, s_i) = E(c_i)$. Una limitación importante de este supuesto es que descarta que la selección pueda depender de la heterogeneidad no observada c_i .

- 1 Sesgo de Selección: “Attrition” y Truncamiento Incidental
 - Cuándo Podemos Ignorar el Sesgo de Selección?
 - Contrastes por Sesgo de Selección
 - Corrección del Sesgo de Selección Muestral: Truncamiento Incidental
 - Procedimiento de Wooldridge
 - Procedimiento de Rochina-Barrachina
 - Corrección del Sesgo de Selección Muestral: Attrition
- 2 Sesgo de Selección y Endogeneidad

Paneles No Balanceados: Inferencia

- Un contraste de hipótesis simple fue desarrollado por Nijman y Verbeek (1992): adicione el indicador de selección rezagado, $s_{i,t-1}$, a la ecuación. Estime el modelo usando efectos fijos (sobre el panel desbalanceado), y haga un contraste t (haciéndolo completamente robusto) para chequear si $s_{i,t-1}$ es estadísticamente relevante.
- Bajo la hipótesis nula, u_{it} no está correlacionado con s_{ir} para todo r , y entonces el indicador de selección en el período temporal anterior no debiera ser relevante en la ecuación del período t . (Incidentalmente, nunca tendría sentido poner s_{it} en la ecuación del período t porque $s_{it} = 1$ para todo i y t en la sub-muestra seleccionada.)

Paneles No Balanceados: Inferencia

- Poner $s_{i,t-1}$ no funciona si $s_{i,t-1}$ es uno siempre que s_{it} es uno porque entonces no hay variación en $s_{i,t-1}$ en la muestra seleccionada. Este es el caso de los problemas de attrition donde, digamos, una persona solo puede aparecer en el período t si el o ella aparecieron en $t - 1$.
- Una alternativa es incluir un adelanto del indicador de selección, $s_{i,t+1}$. Para las observaciones i que están en la muestra todos los períodos, $s_{i,t+1}$ siempre es uno. Pero para los que se van de la muestra (attriters), $s_{i,t+1}$ cambia de uno a cero justo en el período anterior a salirse del panel.

Paneles No Balanceados: Ejemplo Attrition

- Los datos de la Tabla 1 corresponden a producto en millones de kilovatios-horas y el costo total de generación de energía en millones de dólares para diez firmas observadas durante 4 años.

Tabla 1					
		Tiempo			
Firma		t=1	t=2	t=3	t=4
i=1	Costo	3	4	5	6
	Producto	214	419	588	1025
i=2	Costo	4	6	8	n.a.
	Producto	696	811	1640	n.a.
i=3	Costo	19	26	32	n.a.
	Producto	3202	4802	5821	n.a.
i=4	Costo	35	51	61	n.a.
	Producto	5668	7612	10206	n.a.
i=5	Costo	33	40	n.a.	n.a.

Paneles No Balanceados: Ejemplo Attrition

- Tabla 1 (Cont.)

		Tiempo			
Firma		t=1	t=2	t=3	t=4
i=6	Costo	73	99	n.a.	n.a.
	Producto	11796	15551	n.a.	n.a.
i=7	Costo	80	106	n.a.	n.a.
	Producto	11803	15558	n.a.	n.a.
i=8	Costo	95	n.a.	n.a.	n.a.
	Producto	11818	n.a.	n.a.	n.a.
i=9	Costo	116	142	n.a.	n.a.
	Producto	11839	15594	n.a.	n.a.
i=10	Costo	144	n.a.	n.a.	n.a.
	Producto	11867	n.a.	n.a.	n.a.

Paneles No Balanceados: Ejemplo Attrition

- Base de datos para implementar el test de Nijman-Verbeek,

Firma	Tiempo	Costo	Producto	$s(i,t+1)$
1	1	3.154	214	1
1	2	4.271	419	1
1	3	4.584	588	1
2	1	3.859	696	1
2	2	5.535	811	1
2	3	8.127	1640	0
3	1	19.035	3202	1
3	2	26.041	4802	1
3	3	32.444	5821	0
4	1	35.229	5668	1
4	2	51.111	7612	1
4	3	61.045	10206	0
5	1	33.154	6000	1
5	2	40.044	8222	0
6	1	73.05	11796	1

Paneles No Balanceados: Ejemplo Attrition

- En Stata,
gen lncit=ln(cost)
gen lnyit=ln(output)
iis firm
tis time
areg lncit lnyit sit1, absorb(firm) robust
outreg using attrition1, 3aster
type attrition1.out

	lncit
lnyit	0.540
	(3.94)***
s(i,t+1)	-0.116
	(1.97)*
Constant	-1.040
	(0.88)

Paneles No Balanceados: Inferencia

- Para los problemas de truncamiento incidental tiene sentido extender el test de Heckman (1976) al contexto de datos de panel con heterogeneidad no observada. Esto es hecho por Wooldridge (1995). Escribamos la ecuación de interés como

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1, 2, \dots, T \quad (3)$$

► Heckit

- Inicialmente supongamos que y_{it} se observa solo si el indicador de selección, s_{it} , es uno.
- Suponga que, para cada t , s_{it} está determinado por la ecuación probit

$$s_{it} = 1[\eta_0 + \bar{x}_i\eta + x_{it}\delta + v_{it} > 0], \quad v_{it}|x_i \sim \text{Normal}(0, 1) \quad (4)$$

- El mecanismo de selección descrito en la ecuación (4) **no necesita estar correctamente especificado para obtener un buen contraste** (vea Wooldridge, 1995).

Paneles No Balanceados: Inferencia

- Bajo la hipótesis nula del test, en el Supuesto FEUP.1 (a) (con los cambios notacionales obvios), la inversa del cociente de Mills obtenida de la estimación del modelo probit no debiera ser estadísticamente relevante en la ecuación estimada por efectos fijos
- Sea $\hat{\lambda}_{it}$ la inversa del cociente de Mills estimada en la ecuación (4) por **pooled probit** a través de i y t . Entonces, un contraste válido de la hipótesis nula es un estadístico t (robusto ante la presencia de heterocedasticidad y correlación serial) sobre el coeficiente de $\ddot{\lambda}_{it} = \hat{\lambda}_{it} - T_i^{-1} \sum_{r=1}^T s_{ir} \hat{\lambda}_{ir}$ en la estimación de FE usando solo aquellas observaciones para las que $s_{it} = 1$,

$$\ddot{y}_{it} = \ddot{x}_{it}\beta + \rho \ddot{\lambda}_{it} + \text{error}_{it}$$

- Wooldridge (1995) muestra formalmente que la estimación probit del primer paso no afecta la distribución asintótica del estadístico t bajo $H_0 : \rho = 0$ en la última ecuación.

Paneles No Balanceados: Ejemplo Truncamiento Incidenta

- Mismos datos que en la Tabla 1, pero...

		Tiempo			
Firma		t=1	t=2	t=3	t=4
i=1	Costo	3	4	5	6
	Producto	214	419	588	1025
i=2	Costo	4	6	8	n.a.
	Producto	696	811	1640	2506
i=3	Costo	19	26	32	n.a.
	Producto	3202	4802	5821	9275
i=4	Costo	35	51	61	n.a.
	Producto	5668	7612	10206	13702
i=5	Costo	33	40	n.a.	n.a.
	Producto	6000	8222	n.a.	10004
:	:	:	:	:	:

Paneles No Balanceados: Truncamiento Incidental

- **En Stata:** sort firm time
by firm: egen \overline{lnyit} =mean(lnyit)
probit sit lnyit \overline{lnyit}
predict sitf, xb
gen lambdait=normden(sitf)/norm(sitf)
areg Incit lnyit lambdait, absorb(firm) robust
outreg using itrunc, 3aster
type itrunc.out

	Incit
lnyit	0.521
	(4.33)***
$\hat{\lambda}_{it}$	0.140
	(1.66)
Constant	-1.012
	(1.02)
Observations	23
R-squared	0.99

- 1 Sesgo de Selección: “Attrition” y Truncamiento Incidental
 - Cuándo Podemos Ignorar el Sesgo de Selección?
 - Contrastes por Sesgo de Selección
 - Corrección del Sesgo de Selección Muestral: Truncamiento Incidental
 - Procedimiento de Wooldridge
 - Procedimiento de Rochina-Barrachina
 - Corrección del Sesgo de Selección Muestral: Attrition
- 2 Sesgo de Selección y Endogeneidad

Truncamiento Incidental: Estimación

- Corregir el sesgo de selección en el caso de truncamiento incidental requiere mucho más cuidado.
- Desafortunadamente, bajo cualquier supuesto que permita heterogeneidad no observada en la ecuación de selección, adicionar $\hat{\lambda}_{it}$ a la ecuación (3) y usar FE no produce estimadores consistentes (Wooldridge, 1995).
- Para tener un método que funcione necesitamos agregar algunos supuestos de linealidad a los valores esperados de u_{it} y c_i dados x_i y v_{it} .
- **Supuesto FEUP.2:** (a) La ecuación de selección está dada por (4); (b) $E(u_{it}|x_i; v_{it}) = E(u_{it}|v_{it}) = \rho_t v_{it}$, $t = 1; \dots; T$; y (c) $E(c_i|x_i; v_{it}) = L(c_i|1; \bar{x}_i; v_{it})$, donde $L(\cdot|\cdot)$ es el operador proyección lineal.
- El **Supuesto FEUP.2 (b)** es estándar y se sigue del supuesto de normalidad conjunta de (u_{it}, v_{it}) como en Heckman (1976).

Truncamiento Incidental: Estimación

- El **Supuesto FEUP.2 (c)** implica que c_i depende (linealmente) de x_i solo a través del promedio temporal,

$$E(c_i|x_i, v_{it}) = \pi_0 + \bar{x}_i\pi + \phi_tv_{it}.$$

- Tomando esperanzas condicionales en (3),

$$\begin{aligned} E(y_{it}|x_i, v_{it}) &= x_{it}\beta + E(c_i|x_i, v_{it}) + E(u_{it}|x_i, v_{it}) \\ &= \pi_0 + x_{it}\beta + \bar{x}_i\pi + \phi_tv_{it} + \rho_tv_{it} \\ &= \pi_0 + x_{it}\beta + \bar{x}_i\pi + \gamma_tv_{it} \end{aligned}$$

- Condicionando en $s_{it} = 1$,

$$\begin{aligned} E(y_{it}|x_i, s_{it} = 1) &= \pi_0 + x_{it}\beta + \bar{x}_i\pi + \gamma_tE(v_{it}|x_i, s_{it} = 1) \\ &= \pi_0 + x_{it}\beta + \bar{x}_i\pi + \gamma_tE(v_{it}|x_i, v_{it} > -\eta_0 - \bar{x}_i\eta - x_{it}\delta) \\ &= \pi_0 + x_{it}\beta + \bar{x}_i\pi + \gamma_t\lambda_{it}, \end{aligned}$$

Truncamiento Incidental: Estimación

- La última ecuación lleva al siguiente procedimiento,
 - Estime la ecuación (4) por **pooled probit** a través de i y t . Para $s_{it} = 1$ obtenga la inversa del cociente de Mills estimado

$$\hat{\lambda}_{it} = \phi(\hat{\eta}_0 + \bar{x}_i \hat{\eta} + x_{it} \hat{\delta}) / \Phi(\hat{\eta}_0 + \bar{x}_i \hat{\eta} + x_{it} \hat{\delta}) = \phi(h_{it} \hat{\mu}) / \Phi(h_{it} \hat{\mu}).$$

donde $\phi(\cdot)$ y $\Phi(\cdot)$ son la densidad y la distribución acumulada de la normal estándar, respectivamente. $h_{it} \equiv (1 \ \bar{x}_i \ x_{it})$; $\hat{\mu} = (\hat{\eta}_0 \ \hat{\eta}' \ \hat{\delta}')'$.

- Para $s_{it} = 1$ defina el vector $1 \times (1 + 2K + T)$,

$$\hat{w}_{it} = (1, \underbrace{\bar{x}_i}_K, \underbrace{x_{it}}_K, \underbrace{0, \dots, 0, \hat{\lambda}_{it}, 0, \dots, 0}_T),$$

y obtenga $\hat{\theta} = (\hat{\pi}_0, \hat{\pi}', \hat{\beta}', \hat{\gamma}')'$ como el estimador pooled OLS en

$$y_{it} = \hat{w}_{it} \theta + \text{error}_{it}, \quad s_{it} = 1.$$

- Esto da,

$$\hat{\theta} = \left(\sum_{i=1}^N \sum_{t=1}^T s_{it} \hat{w}_{it}' \hat{w}_{it} \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T s_{it} \hat{w}_{it}' y_{it} \right).$$

- $\hat{\theta}$ será consistente y \sqrt{N} -asintóticamente normal. Sin embargo, la inferencia no es estándar debido a la presencia de $\hat{\lambda}_{it}$ en el segundo paso de la estimación.

Truncamiento Incidental: Inferencia

- Estime la varianza asintótica de $\hat{\theta}$, $\text{Avar}(\hat{\theta})$ como sigue:
 1. Defina los residuos de pooled OLS $\hat{e}_{it} \equiv y_{it} - \hat{w}_{it}\hat{\theta}$, $s_{it} = 1$.
 2. Defina la matriz

$$\hat{D} \equiv N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \hat{w}_{it}' \hat{\theta}' G_{it}$$

donde,

$$G_{it} = \begin{pmatrix} 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & Z_{it} & 0 & \cdots & 0 \end{pmatrix}$$

es $(1 + 2K + T) \times T(1 + 2K)$. Cada cero en la primera fila (bloque) de G_{it} es una matriz de $(1 + 2K) \times (1 + 2K)$ y cada cero en la segunda fila (bloque) es una matriz de $T \times (1 + 2K)$. La matriz Z_{it} está en la t -ésima columna (bloque) de la matriz.

Truncamiento Incidental: Inferencia

- Continuación

2.(Cont.) La matriz, $T \times (1 + 2K)$, Z_{it} se define como

$$Z_{it} = (0' 0' \dots 0' (\dot{\lambda}_{it} h_{it})' 0' \dots 0')'$$

donde $\dot{\lambda}_{it}$ es la derivada de λ_{it} evaluada en $h_{it}\hat{\mu}$. Cada cero en Z_{it} es $1 \times (1 + 2K)$.

3. Para cada i , \hat{r}_i es el vector $(1 + 2K) \times 1$ igual a menos la inversa del Hesiano promedio por la función score del logaritmo de la función de verosimilitud del probit estimada para la observación i . [► Go](#)

4. Obtenga

$$\hat{A} = N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \hat{W}_{it}' \hat{W}_{it}$$

Truncamiento Incidental: Inferencia

(cont.) Para cada i defina,

$$\hat{q}_i = \sum_{t=1}^T s_{it} \hat{w}'_{it} \hat{e}_{it} \text{ y } \hat{p}_i = \hat{q}_i - \hat{D} \hat{r}_i$$


- 5. Defina,

$$\hat{B} = N^{-1} \sum_{i=1}^N \sum_{t=1}^T \hat{p}_i \hat{p}'_i$$

Entonces,

$$\widehat{\text{Avar}}(\hat{\theta}) = \hat{A}^{-1} \hat{B} \hat{A}^{-1} / N$$

Paneles No Balanceados: Truncamiento Incidental

- Un segundo caso que a veces se encuentra en la práctica es aquel en el que **la variable de selección se observa en forma parcial.** 
- En economía uno de los casos en donde esta situación aparece es en estudios del mercado de trabajo donde las ecuaciones son las de oferta de trabajo y las de salario.
- En este caso la ecuación de interés es:

$$y_{jt} = x_{jt}^1 \beta + c_j + u_{jt}, \quad t = 1, 2, \dots, T \quad (5)$$

- La variable y_{jt} se observa siempre que h_{jt} se observa. Para cada $t = 1, 2, \dots, T$, definamos la siguiente variable censurada:

$h_{jt} = \max(0, h_{jt}^*)$, se observa, y

$$h_{jt}^* = \delta_{t0} + x_{j1}\delta_{t1} + \dots + x_{jT}\delta_{tT} + v_{jt}, \quad (6)$$

Paneles No Balanceados: Truncamiento Incidental

- x_{jt} denota el conjunto total de variables explicativas en el período t y x_{jt}^1 es un subconjunto de esas variables.
- El mecanismo de selección descrito por la ecuación (6) puede ser visto como la forma reducida de la ecuación de selección.
- Los supuestos necesarios para derivar la generalización del procedimiento de Heckman en este caso son.
- **Supuesto W.1:**
 - (a) La ecuación de selección está dada por (6) con todos sus supuestos;
 - (b) $E(u_{jt} \mid X_j, v_{jt}) = E(u_{jt} \mid v_{jt}) = \rho_t v_{jt}$, $t = 1, 2, \dots, T$; y
 - (c) $E(c_j \mid X_j, v_{jt}) = L(c_j \mid X_j, v_{jt})$.

Paneles No Balanceados: Truncamiento Incidental

- La estrategia utilizada para estimar en forma consistente en este caso es:
- (1) Para cada t , estime la ecuación:

$$h_{jt} = \max(0, X_j \delta_t + v_{jt})$$

usando un modelo Tobit estándar.

- Recuerde que en este set up $X_j = (1, x_{j1}, x_{j2}, \dots, x_{jt})$ y $\delta_t = (\delta_{t0}, \delta_{t1}, \dots, \delta_{tT})$.
Para $s_{jt} = 1$, defina

$$\hat{v}_{jt} = h_{jt} - X_j \hat{\delta}_t$$

- y el vector $1 \times (1 + TK + K + T)$:

$$\hat{w}_{jt} = (1, x_{j1}, \dots, x_{jT}, 0, \dots, 0, \hat{v}_{jt}, 0, \dots, 0).$$

Paneles No Balanceados: Truncamiento Incidental

- (2) Aplique POLS a:

$$y_{jt} = \hat{w}_{jt} \begin{pmatrix} \hat{\pi} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} + error_{jt}; s_{jt} = 1$$

- Lo que da:

$$\hat{\theta} = \begin{pmatrix} \hat{\pi} \\ \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \left(\sum_{j=1}^N \sum_{t=1}^T s_{jt} \hat{w}_{jt}' \hat{w}_{jt} \right)^{-1} \left(\sum_{j=1}^N \sum_{t=1}^T s_{jt} \hat{w}_{jt}' y_{jt} \right)$$

Paneles No Balanceados: Truncamiento Incidental

- (3) Estime la varianza asintótica de los estimadores, $Av\hat{ar}(\hat{\theta})$, como sigue:
Primero defina los residuos, $\hat{e}_{jt} \equiv y_{jt} - \hat{w}_{jt}\hat{\theta}$, donde $s_{jt} = 1$.
- Defina la matriz: $\hat{D} \equiv N^{-1} \sum_{j=1}^N \sum_{t=1}^T s_{jt} \hat{w}_{jt}' \hat{\theta}' G_{jt}$.
- Donde $G_{jt} = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & Z_{jt} & 0 & \dots & 0 \end{pmatrix}$ es una matriz de dimensión: $(1 + TK + K + T) \times T(1 + TK)$.
- Cada cero en la primera fila de G_{jt} es de dimensión $(1 + TK + K) \times (1 + TK)$, y cada cero en la segunda fila de la matriz es de dimensión $T \times (1 + TK)$. La matriz Z_{jt} está ubicada a partir de la columna t-ésima y es de dimensión $T \times (1 + TK)$.

$$Z_{jt} = (0', 0', \dots, 0', -X_j, 0', \dots, 0')'$$

Paneles No Balanceados: Truncamiento Incidental

- Cada cero en Z_{jt} es de dimensión $1 \times (1 + TK)$ y el vector $1 \times (1 + TK)$, $-X_j$, está ubicado en la t -ésima fila.
- Para cada t , defina \hat{r}_{jt} como el vector $(1 + TK) \times 1$ igual a menos la inversa de la matriz Hesiana promedio estimada multiplicada por la derivada de la función del logaritmo de verosimilitud del modelo Tobit para la observación j (recuerde eliminar el último elemento de este vector porque la varianza no aparece en \hat{v}_{jt}). [▶ Go](#)
- Construya el vector \hat{r}_j de dimensión $T(1 + TK) \times 1$ “stacking” $\{\hat{r}_{j1}, \hat{r}_{j2}, \dots, \hat{r}_{jT}\}$.

Paneles No Balanceados: Truncamiento Incidental

- La matriz \hat{A} viene dada por: $\hat{A} \equiv N^{-1} \sum_{j=1}^N \sum_{t=1}^T s_{jt} \hat{w}'_{jt} \hat{w}_{jt}$
- Además, para cada $j = 1, 2, \dots, N$ defina: $\hat{q}_j = \sum_{t=1}^T s_{jt} \hat{w}'_{jt} \hat{e}_{jt}$ y $\hat{p}_j = \hat{q}_j - \hat{D} \hat{r}_j$.
- Un estimador consistente de B es: $\hat{B} \equiv N^{-1} \sum_{j=1}^N \hat{p}_j \hat{p}'_j$
- Finalmente la varianza asintótica de los estimadores del modelo viene dada por:

$$Avâr \left(\hat{\theta} \right) = \hat{A}^{-1} \hat{B} \hat{A}^{-1} / N$$

- Y los desvíos estándar asintóticos se obtienen con la raíz cuadrada de los elementos de la diagonal de esta matriz.
- **Remark 2.** Como ocurría con el caso anterior, la corrección propuesta por Wooldridge requiere que se cumpla el supuesto de exogeneidad estricta de los regresores (condicional a los efectos no observables).

Paneles No Balanceados: Truncamiento Incidental

- Existe un estimador alternativo al de Wooldridge para el caso en que el mecanismo de selección viene dado por una variable binaria.
- Este estimador, sugerido por Rochina-Barrachina (1999) (RB a partir de ahora) difiere del estimador de Wooldridge en que permite que la media condicional de los efectos no observables de la ecuación de interés sea desconocida.
- Para levantar el supuesto W.1(c) de Wooldridge, RB impone el supuesto de que la distribución conjunta de los errores de la ecuación de interés en primeras diferencias y los errores de las dos ecuaciones de selección (correspondientes a los dos períodos de las diferencias finitas), condicional al vector completo de variables explicativas estrictamente exógenas, es normal.

Paneles No Balanceados: Truncamiento Incidental

- RB desarrolla su estimador en el contexto de un panel en donde el número de observaciones de corte transversal es grande y las propiedades asintóticas del estimador son válidas con $N \rightarrow \infty$.
- El desarrollo se basa en dos períodos temporales, $T = 2$, y es como lo estudiaremos aquí.
- La idea básica del estimador es, primero, eliminar los efectos no observables de la ecuación de interés tomando diferencias.

Paneles No Balanceados: Truncamiento Incidental

- Después, condicionando en que el resultado del proceso de selección sea uno en los dos períodos temporales, construir la ecuación a estimar en el segundo paso. [▶ Go](#)
- Esta ecuación contiene dos términos de corrección por sesgo de selección muestral
- La ecuación de interés es:

$$y_{jt} = x_{jt}^1 \beta + c_j + u_{jt}, \quad t = 1, 2, \dots, T \quad (7)$$

- RB plantea la siguiente ecuación estructural para el mecanismo de selección:

$$h_{jt}^* = \xi_j + x_{jt} \delta + a_{jt},$$

$$\text{con } s_{jt} = 1 [h_{jt}^* \geq 0],$$

Paneles No Balanceados: Truncamiento Incidental

- Donde c_j y ξ_j son efectos específicos de corte transversal no observables que probablemente estén correlacionados con las variables explicativas observables.
- u_{jt} y a_{jt} son errores idiosincráticos no necesariamente independientes uno del otro.
- Para obtener estimadores consistentes en (7) utilizando solo las observaciones de la muestra seleccionada necesitamos la siguiente condición:

$$E(u_{jt} - u_{js} \mid X_j, s_{jt} = s_{js} = 1) = 0, s \neq t \quad (8)$$

- Esta condición está expresada en forma diferente a la que utilizamos cuando vimos la estimación usando diferencias finitas.

Paneles No Balanceados: Truncamiento Incidental

- Note que en la condición anterior, a diferencia de lo que hacíamos con FD, no necesariamente necesitamos el operador diferencias finitas de orden uno. De hecho, cualquier diferencia temporal entre dos observaciones del mismo corte transversal eliminará el componente c_j .
- En particular, esta condición nos permitirá trabajar con aquellas observaciones que tengan $s_{jt} = s_{js} = 1 (t \neq s)$.

Paneles No Balanceados: Truncamiento Incidental

- Como mencionamos anteriormente, la condición (8) en general será diferente de cero debido al sesgo de selección muestral.
- Los supuestos necesarios para derivar la corrección propuesta por RB son:
- **Supuesto RB.1** (a) ξ_j está correlacionado con X_j , a través de la siguiente especificación:
$$\xi_j = \eta_0 + x_{j1}\eta_1 + \dots + x_{jT}\eta_T + \alpha_j$$
- (b) los errores de la ecuación de selección tienen distribución normal, $v_{jt} = a_{jt} + \xi_j \sim N(0, \sigma_t^2)$.
- (c) los errores $[(u_{jt} - u_{js}), v_{jt}, v_{js}]$ tienen distribución conjunta trivariada normal condicional a X_j .

Paneles No Balanceados: Truncamiento Incidental

- Bajo el **supuesto RB.1**, la forma funcional del término de sesgo de selección se puede derivar de la generalización del Teorema 20.4 de Greene para el caso de la distribución normal multivariante truncada.
- La esperanza condicional (8) puede escribirse como:
$$E(u_{jt} - u_{js} \mid X_j, v_{jt} \geq -H_{jt}, v_{js} \geq -H_{js}) = \sigma_{(u_j - u_s)(v_t/\sigma_t)} \lambda_{jts} + \sigma_{(u_j - u_s)(v_s/\sigma_s)} \lambda_{jst},$$
- Donde $H_{j\tau} = X_{j\tau}\delta + E(\xi_j \mid X_j)$ para $\tau = t, s$, es la forma reducida de los indicadores de selección para los períodos t y s .
- Además, $\lambda_{jts} = \phi(M_{jt})\Phi(M_{jts}^*) / \Phi_2(M_{jt}, M_{js}, \rho_{ts})$, y
 $\lambda_{jst} = \phi(M_{js})\Phi(M_{jst}^*) / \Phi_2(M_{jt}, M_{js}, \rho_{ts})$.
- Donde, $M_{jt} = \frac{H_{jt}}{\sigma_t}$, $M_{js} = \frac{H_{js}}{\sigma_s}$, $M_{jts}^* = (M_{js} - \rho_{ts}M_{jt})/(1 - \rho_{ts}^2)^{1/2}$ y
 $M_{jst}^* = (M_{jt} - \rho_{ts}M_{js})/(1 - \rho_{ts}^2)^{1/2}$

Paneles No Balanceados: Truncamiento Incidental

- y $\rho_{ts} = \rho(v_t/\sigma_t)(v_s/\sigma_s)$ es el coeficiente de correlación entre los errores del proceso de selección.
- Como antes $\phi(\cdot)$ es la función de densidad y $\Phi(\cdot)$ y $\Phi_2(\cdot)$ son las funciones acumuladas de la distribución normal estándar univariante y bivalente, respectivamente
- Por lo tanto, la ecuación a estimar queda:

$$y_{jt} - y_{js} = (x_{jt}^1 - x_{js}^1) \beta + \ell_{ts} \lambda(M_{jt}, M_{js}, \rho_{ts}) + \ell_{st} \lambda(M_{js}, M_{jt}, \rho_{ts}) + e_{jts} \quad (9)$$

- Donde $e_{jts} \equiv (u_{jt} - u_{js}) - [\ell_{ts} \lambda(M_{jt}, M_{js}, \rho_{ts}) + \ell_{st} \lambda(M_{js}, M_{jt}, \rho_{ts})]$ Es un nuevo término de error que por construcción satisface:

$$E(u_{jt} - u_{js} \mid X_j, v_{jt} \geq -H_{jt}, v_{js} \geq -H_{js}) = 0$$

- Con estas condiciones la solución del problema es inmediata.

Paneles No Balanceados: Truncamiento Incidental

- Si podemos estimar consistentemente, λ_{jts} y λ_{jst} , POLS en (9) puede utilizarse para obtener estimaciones consistentes de β , ℓ_{ts} y ℓ_{st} .
- En la práctica, para construir estimaciones consistentes de los términos λ_{jts} y λ_{jst} , los coeficientes (δ, ρ_{ts}) se determinan conjuntamente usando un modelo Probit bivariado para cada combinación de períodos temporales. El segundo paso, consiste en aplicar POLS a (9)

► Go

Paneles No Balanceados: Truncamiento Incidental

- Una de las complicaciones del estimador de RB es que todo el análisis está basado en $T = 2$. Si $T > 2$, entonces en el primer paso se pueden estimar $\binom{T}{2}$ modelos Probit bivariados, cada uno de ellos basado en una combinación diferente de dos períodos temporales.
- Una vez que las estimaciones de los términos que corrigen el sesgo de selección muestral se incluyen en (9), esta ecuación puede ser estimada para cada combinación de ondas del panel (t, s) , $t \neq s$, lo que da un total de $\binom{T}{2}$ pares para un panel de largo T .

Paneles No Balanceados: Truncamiento Incidental

- Por lo tanto para $T > 2$, hay que utilizar un procedimiento que combine todas estas estimaciones en una sola.
- RB sugieren un procedimiento de mínima distancia (i.e. GMM) con su correspondiente matriz ponderadora.
- Para construir la matriz ponderadora se requiere la estimación de la matriz de varianzas y covarianzas de los estimadores para los diferentes períodos temporales.

- 1 Sesgo de Selección: “Attrition” y Truncamiento Incidental
 - Cuándo Podemos Ignorar el Sesgo de Selección?
 - Contrastes por Sesgo de Selección
 - Corrección del Sesgo de Selección Muestral: Truncamiento Incidental
 - Procedimiento de Wooldridge
 - Procedimiento de Rochina-Barrachina
 - Corrección del Sesgo de Selección Muestral: Attrition

- 2 Sesgo de Selección y Endogeneidad

Attrition: Marco de Análisis

- Attrition general, donde las unidades de corte transversal re-ingresan en la muestra después de dejarla, es complicado.
- Vamos a analizar un caso especial.
- En $t = 1$ se obtiene una muestra aleatoria de la población relevante.
- En $t = 2$ algunas unidades de corte transversal eligen salirse del panel por razones que no son enteramente aleatorias.
- Asumimos que, una vez que la persona se sale del panel, el o ella sale para siempre: attrition es un estado **absorvente**.
- Cualquier panel con attrition puede ser construido de esta manera, ignorando cualquier observación subsiguiente de las unidades de corte transversal que ya han salido de la muestra.

Attrition: Corrección

- Considere el siguiente modelo.

$$y_{jt} = x_{jt}\beta + c_j + u_{jt}, \quad (10)$$

donde asumimos que (x_{jt}, y_{jt}) se observa para todo j cuando $t = 1$.

- Hagamos que s_{jt} denote el indicador de selección para cada período temporal, donde $s_{jt} = 1$ si (x_{jt}, y_{jt}) se observan.
- Como ignoramos las unidades una vez que han salido de la muestra, $s_{jt} = 1$ implica $s_{jr} = 1$ para $r < t$.
- Tomemos diferencias finitas de primer orden.
-

$$\Delta y_{jt} = \Delta x_{jt}\beta + \Delta u_{jt}, \quad t = 2, \dots, T \quad (11)$$

Attrition: Corrección

- Condicional en $s_{jt-1} = 1$, escribamos la forma reducida de la ecuación de selección para $t \geq 2$ como

$$s_{jt} = 1[w_{jt}\delta_t + v_{jt} > 0], \quad v_{jt}|\{w_{jt}, s_{jt-1} = 1\} \sim \text{Normal}(0, 1) \quad (12)$$

donde w_{jt} debe contener variables observadas en t para todas las unidades con $s_{jt-1} = 1$.

- Buenos candidatos para w_{jt} incluyen las variables en x_{jt-1} y cualquier variable in x_{jt} que sea observada en t cuando $s_{jt-1} = 1$ (por ejemplo, si x_{jt} contiene rezagos de variables o una variable como edad).
- Sin embargo, como y_{jt-1} está correlacionada con u_{jt-1} no debería incluirse en w_{jt} .

Attrition: Corrección

- Si las x_{jt} son estrictamente exógenas y la selección no depende de Δx_{jt} una vez que se controla por w_{jt} , un supuesto razonable (bajo normalidad conjunta de Δu_{jt} y v_{jt} es

$$E(\Delta u_{jt} | \Delta x_{jt}, w_{jt}, v_{jt}, s_{jt-1} = 1) = E(\Delta u_{jt} | v_{jt}) = \rho_t v_{jt} \quad (13)$$

- Entonces,

$$E(\Delta y_{jt} | \Delta x_{jt}, w_{jt}, s_{jt} = 1) = \Delta x_{jt} \beta + \rho_t \lambda(w_{jt} \delta_t), \quad t = 2, \dots, T \quad (14)$$

- Note como, porque $s_{jt-1} = 1$ cuando $s_{jt} = 1$, no necesitamos condicionar en s_{jt-1} .

Attrition: Corrección

- Se sigue de la última ecuación que **pooled OLS** de Δy_{jt} sobre Δx_{jt} , $d2_t \hat{\lambda}_{jt}$, \dots , $dT_t \hat{\lambda}_{jt}$, $t = 2, \dots, T$, donde los $\hat{\lambda}_{jt}$ vienen de los $T - 1$ probits de corte transversal de la ecuación (12), es consistente para β y ρ_t .
- Un contraste conjunto (completamente robusto) de $H_0 : \rho_t = 0$, $t = 2, \dots, T$ es un test simple por la presencia de sesgo de selección por attrition.
- Hay dos problemas potenciales con este enfoque. 1. La primera igualdad en (13) es restrictiva porque significa que x_{jt} no afecta la attrition una vez que se controla por los elementos de w_{jt} . 2. Asumimos exogeneidad estricta de x_{jt} . Ambas restricciones se pueden relajar con un procedimiento de IV.

Estimación en Presencia de Sesgo de Selección y Endogeneidad: Pooled 2SLS (Wooldridge & Semykina, 2010)

- Considere el siguiente modelo,

$$y_{it} = x_{it}\beta + v_{it} \quad (15)$$

donde x_{it} es un vector $1 \times K$ que contiene variables exógenas y **endógenas**, β es un vector $K \times 1$ de parámetros y v_{it} es el error compuesto.

- Adicionalmente, asumamos que existe un vector $L \times 1$ de instrumentos ($L \geq K$), z_{it} , tal que el supuesto de **exogeneidad contemporánea** se cumple para todas las variables en z_{it} : $E(v_{it}|z_{it}) = 0$, $t = 1, 2, \dots, T$.
- Asumimos también que: (i) los vectores x_{it} y z_{it} contienen una constante; (ii) los instrumentos están suficientemente correlacionados con las variables explicativas en el análogo poblacional de (15); (iii) z_{it} incluye todas las variables de x_{it} que son exógenas.

Estimación en Presencia de Sesgo de Selección y Endogeneidad: Pooled 2SLS

- Ahora introduzcamos **s sesgo de selección (truncamiento incidental)** en el modelo.
- Sea s_{it} **un indicador de selección** que vale **uno** si (y_{it}, x_{it}, z_{it}) se observa y vale **cero** en otro caso.
- Entonces,

$$\begin{aligned}\hat{\beta}_{2SLS} = & \left[\left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} x'_{it} z_{it} \right) \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} z'_{it} z_{it} \right)^{-1} \right. \\ & \times \left. \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} z'_{it} x_{it} \right) \right]^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} x'_{it} z_{it} \right) \\ & \times \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} z'_{it} z_{it} \right)^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} z'_{it} y_{it} \right) \quad (16)\end{aligned}$$

Estimación en Presencia de Sesgo de Selección y Endogeneidad: Pooled 2SLS

- La ecuación (16) se puede escribir como,

$$\begin{aligned}\hat{\beta}_{2SLS} = & \beta + \left[\left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} x'_{it} z_{it} \right) \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} z'_{it} z_{it} \right)^{-1} \right. \\ & \times \left. \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} z'_{it} x_{it} \right) \right]^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} x'_{it} z_{it} \right) \\ & \times \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} z'_{it} z_{it} \right)^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} z'_{it} v_{it} \right) \quad (17)\end{aligned}$$

- Para T fijo y $N \rightarrow \infty$ las siguientes condiciones establecen la **consistencia del estimador Pooled 2SLS**.

Estimación en Presencia de Sesgo de Selección y Endogeneidad: Pooled 2SLS

- **Supuesto P2SLS.1:** (i) (y_{it}, x_{it}, z_{it}) se observa cuando $s_{it} = 1$; (ii) $E(v_{it}|z_{it}, s_{it}) = 0$, $t = 1, 2, \dots, T$; (iii) rango $E\left(\sum_{t=1}^T s_{it} z'_{it} x_{it}\right) = K$; (iv) rango $E\left(\sum_{t=1}^T s_{it} z'_{it} z_{it}\right) = L$.
- El **Supuesto P2SLS.1 (iii)** es la condición de rango importante (sobre la subpoblación observada) y requiere que haya suficientes instrumentos ($L \geq K$) y que estén correlacionados con x_{it} .
- El **Supuesto P2SLS.1 (ii)** establece el sentido en el que **la selección se asume exógena en (17)**. Requiere que v_{it} sea independiente en media (condicionalmente) de z_{it} y el mecanismo de selección en t .
- **Supuesto P2SLS.2:** Bajo el **Supuesto P2SLS.1** y condiciones de regularidad estándar el **estimador Pooled 2SLS es consistente y asintóticamente normal**.

Estimación en Presencia de Sesgo de Selección y Endogeneidad: Pooled 2SLS

- **Remark 1:** Note que el **Supuesto P2SLS.1 (ii)** no restringe la relación entre v_{it} y s_{ir} , para $r \neq t$. En otras palabras, el mecanismo de selección se asume exógeno en forma contemporánea pero no en forma estricta.
- **Remark 2:** El **Supuesto P2SLS.1** no impone ninguna relación acerca de la naturaleza de la endogeneidad de los elementos de x_{it} (i.e. podemos tener variables binarias endógenas en este contexto).

Estimación en Presencia de Sesgo de Selección y Endogeneidad: FE-2SLS

- Como mencionamos anteriormente, en muchas aplicaciones de datos de panel queremos introducir en el modelo heterogeneidad no observada que esté correlacionada con las variables explicativas observadas y, en este caso particular, también con las variables instrumentales.
- Considere el modelo de componentes no observados,

$$y_{it} = x_{it}\beta + c_i + u_{it} \quad (18)$$

donde c_i es el efecto individual no observado y u_{it} es el error idiosincrático.

- Supuestos: (i) correlación arbitraria entre la heterogeneidad no observada y las variables explicativas observadas; (ii) permitimos que algunos elementos de x_{it} estén correlacionados con el error idiosincrático u_{it} ; (iii) existen instrumentos z_{it} que son estrictamente exógenos condicionados en c_i .

Estimación en Presencia de Sesgo de Selección y Endogeneidad: FE-2SLS

- Estos supuestos permiten correlación entre z_{it} y c_i pero requieren que los z_{it} no estén correlacionados con $\{u_{ir} : r = 1, 2, \dots, T\}$.
- Como el modelo de efectos fijos requiere de alguna transformación para eliminar los c_i , asumimos que **todas las variables en x_{it} y z_{it} varían en el tiempo**.
- **Bajo que condiciones ignorar la selección produce un estimador consistente?**
- definamos para cada i y t : $\ddot{x}_{it} = x_{it} - (1/T_i) \sum_{r=1}^T s_{ir} x_{ir}$;
 $\ddot{z}_{it} = z_{it} - (1/T_i) \sum_{r=1}^T s_{ir} z_{ir}$; $\ddot{y}_{it} = y_{it} - (1/T_i) \sum_{r=1}^T s_{ir} y_{ir}$ and $T_i = \sum_{r=1}^T s_{ir}$.
- Entonces **el estimador de 2SLS de efectos fijos (FE-2SLS)** es:

Estimación en Presencia de Sesgo de Selección y Endogeneidad: FE-2SLS

$$\begin{aligned}\hat{\beta}_{FE-2SLS} = & \left[\left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{x}'_{it} \ddot{z}_{it} \right) \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{z}_{it} \right)^{-1} \right. \\ & \times \left. \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{x}_{it} \right) \right]^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{x}'_{it} \ddot{z}_{it} \right) \\ & \times \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{z}_{it} \right)^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{y}_{it} \right) \quad (19)\end{aligned}$$

- Usando álgebra,

Estimación en Presencia de Sesgo de Selección y Endogeneidad: FE-2SLS

$$\begin{aligned}\hat{\beta}_{FE-2SLS} &= \beta + \left[\left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{x}'_{it} \ddot{z}_{it} \right) \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{z}_{it} \right)^{-1} \right. \\ &\quad \times \left. \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{x}_{it} \right) \right]^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{x}'_{it} \ddot{z}_{it} \right) \\ &\quad \times \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{z}_{it} \right)^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it} \ddot{z}'_{it} u_{it} \right) \quad (20)\end{aligned}$$

- Denote por $z_i = (z_{i1}, \dots, z_{iT})$ y $s_i = (s_{i1}, \dots, s_{iT})$. Para que el estimador FE-2SLS sea consistente en paneles no balanceados necesitamos los siguientes supuestos:

Estimación en Presencia de Sesgo de Selección y Endogeneidad: FE-2SLS

- **Supuesto FE2SLS.1:** (i) (y_{it}, x_{it}, z_{it}) se observa cuando $s_{it} = 1$;
(ii) $E(u_{it}|z_i, s_i, c_i) = 0, t = 1, 2, \dots, T$;
(iii) $\text{rango } E\left(\sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{x}_{it}\right) = K$;
(iv) $\text{rango } E\left(\sum_{t=1}^T s_{it} \ddot{z}'_{it} \ddot{z}_{it}\right) = L$.
- Asumiendo que hay suficientes instrumentos que varíen en el tiempo el **Supuesto FE2SLS.1 (ii)** es el supuesto crítico. Usando la ley de expectativas iteradas, el **Supuesto FE2SLS.1 (ii)** garantiza que $E(\sum_{t=1}^T s_{it} \ddot{z}'_{it} u_{it}) = 0$ y el último término de la ecuación (20) converge en probabilidad a cero cuando $N \rightarrow \infty$.
- El **Supuesto FE2SLS.1 (ii)** siempre se satisface si los z_{it} son estrictamente exógenos, condicional en c_i , y el mecanismo de selección es completamente aleatorio, tal que s_i es independiente de (u_{it}, z_i, c_i) en todos los períodos

Estimación en Presencia de Sesgo de Selección y Endogeneidad: FE-2SLS

- Permitir correlación arbitraria entre s_{it} y c_i es lo que hace atractivo al modelo de efectos fijos en paneles no balanceados cuando se sospecha que la propensión a dejar el panel (attrition) o seleccionarse fuera del mismo (truncamiento incidental) está relacionado con la heterogeneidad no observada.
- El **Supuesto FE2SLS.1**: sugiere contrastes simples de adición de variables para detectar sesgo de selección.
- Como el **Supuesto FE2SLS.1 (ii)** implica que u_{it} no está correlacionado con s_{ir} para todo t y r , se pueden adicionar funciones, que varíen en el tiempo, de los indicadores de selección y obtener tests t o de Wald.
- En el espíritu de Nijman y Verbeek (1992), se pueden adicionar $s_{i,t-1}$ o $s_{i,t+1}$ a (18) y contrastar por su significancia estadística.

Contraste por Sesgo de Selección bajo Truncamiento Incidental

- El objetivo es contrastar si existe sesgo de selección en la ecuación de interés (i.e. si existe correlación entre el mecanismo de selección y el error idiosincrático).
- Cambiando solo un poco la notación anterior considere la siguiente ecuación de interés:

$$y_{it1} = x_{it}\beta_1 + c_{i1} + u_{it1}, \quad t = 1, \dots, T \quad (21)$$

donde x_{it} es un vector $1 \times K$ que contiene variables exógenas y **endógenas**, β_1 es un vector $K \times 1$ de parámetros y c_{i1} es la heterogeneidad no observada y u_{it1} es el error idiosincrático.

- Adicionalmente, asumimos que existe un vector $L \times 1$ de instrumentos ($L \geq K$), z_{it} , estrictamente exógenos, condicional en c_i .

Contraste por Sesgo de Selección bajo Truncamiento Incidental

- Asumimos que las variables instrumentales z_{it} siempre se observan mientras que (y_{it1}, x_{it1}) solo se observan si el indicador de selección, ahora llamado s_{it2} , vale uno.
- Definamos la variable latente s_{it2}^* ,

$$s_{it2}^* = z_{it}\delta_2 + c_{i2} + u_{it2}, \quad t = 1, \dots, T \quad (22)$$

donde c_{i2} es la heterogeneidad no observada y u_{it2} es un error idiosincrático.

- El mecanismo de selección s_{it2} se genera como

$$s_{it2} = 1[s_{it2}^* > 0] = 1[z_{it}\delta_2 + c_{i2} + u_{it2} > 0], \quad (23)$$

con $1[\cdot]$ es la función indicadora.

- Para derivar el test asumimos que

$$u_{it2}|z_i, c_{i2} \sim \text{Normal}(0, 1), \quad t = 1, \dots, T \quad (24)$$

tal que s_{it2} sigue un modelo Probit de efectos fijos.

Contraste por Sesgo de Selección bajo Truncamiento Incidental

- Igual que en el caso de los modelos sin variables endógenas la heterogeneidad no observada se relaciona con las variables exógenas, z_i , usando el supuesto de Mundlak (1978),

$$c_{i2} = \bar{z}_i \xi_2 + a_{i2}, \quad (25)$$

$$a_{i2}|z_i \sim \text{Normal}(0, \tau_2^2), \quad t = 1, \dots, T \quad (26)$$

que asume que la correlación entre c_{i2} y z_i actúa solo a través de de las medias temporales de las variables exógenas mientras que la parte remanente del componente no observado, a_{i2} , es independiente de z_i .

- Combinando las ecuaciones (22) a (26) el indicador de selección se puede escribir como,

$$\begin{aligned} s_{it2} &= 1[z_{it}\delta_2 + \bar{z}_i\xi_2 + a_{i2} + u_{it2} = z_{it}\delta_2 + \bar{z}_i\xi_2 + v_{it2} > 0], \\ v_{it2}|z_i &\sim \text{Normal}(0, 1 + \tau_2^2), \quad t = 1, \dots, T \end{aligned} \quad (27)$$

Contraste por Sesgo de Selección bajo Truncamiento Incidental

- El modelo de la ecuación (27) es más restrictivo de lo necesario. Uno puede permitir que los coeficientes del modelo de selección varíen en el tiempo,

$$\begin{aligned} s_{it2} &= 1[z_{it}\delta_2 + \bar{z}_i\xi_{2t} + a_{i2} + u_{it2} = z_{it}\delta_2 + \bar{z}_i\xi_{2t} + v_{it2} > 0], \\ v_{it2}|z_i &\sim \text{Normal}(0, 1), \quad t = 1, \dots, T \end{aligned} \quad (28)$$

- Supuesto FE2SLS.2:** (u_{it1}, v_{i2}) es independiente de (z_i, c_{i1}) y (u_{it1}, v_{it2}) es independiente de $(v_{i12}, \dots, v_{i,t-1,2}, v_{i,t+1,2}, \dots, v_{i,T,2})$. Entonces, si $E(u_{it1}|v_{it2})$ es lineal,

$$E(u_{it1}|z_i, c_{i1}, v_{i2}) = E(u_{it1}|v_{i2}) = E(u_{it1}|v_{it2}) = \rho_1 v_{it2}, \quad t = 1, \dots, T \quad (29)$$

donde, por ahora ρ_1 es constante en el tiempo.

- Bajo el supuesto anterior:

$$E(u_{it1}|z_i, c_{i1}, s_{i2}) = \rho_1 E(v_{it2}|z_i, c_{i1}, s_{i2}) = \rho_1 E(v_{it2}|z_i, s_{it2}), \quad t = 1, \dots, T \quad (30)$$

Contraste por Sesgo de Selección bajo Truncamiento Incidental

- Tomando esperanzas condicionales en la ecuación de interes (21) tenemos,

$$\begin{aligned} E(y_{it1}|z_i, c_{i1}, s_{i2}) &= x_{it}\beta_1 + c_{i1} + E(u_{it1}|z_i, c_{i1}, s_{i2}) \\ &= x_{it}\beta_1 + c_{i1} + \rho_1 E(v_{it2}|z_i, s_{it2}) \implies \end{aligned}$$

$$y_{it1} = x_{it}\beta_1 + c_{i1} + \rho_1 E(v_{it2}|z_i, s_{it2}) + e_{it1}, \quad t = 1, \dots, T \quad (31)$$

donde por construcción $E(e_{it1}|z_i, c_{i1}, s_{i2}) = 0$, $t = 1, \dots, T$.

- Si conociéramos $E(v_{it2}|z_i, s_{it2})$, se podría contrastar por sesgo de selección contrastando $H_0 : \rho_1 = 0$ en (31) usando la estimación de FE-2SLS.
- Como solo estaríamos utilizando las observaciones para las que $s_{it2} = 1$ solo necesitamos conocer $E(v_{it2}|z_i, s_{it2} = 1)$.

Contraste por Sesgo de Selección bajo Truncamiento Incidental

- Utilizando el teorema 20.4 de Greene para los momentos de una normal bivalente truncada obtenemos el cálculo Probit usual,

$$E(v_{it2}|z_i, s_{it2} = 1) = \lambda(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2}), \quad t = 1, \dots, T \quad (32)$$

donde $\lambda(\cdot)$ denota la inversa del cociente de Mills.

- Este desarrollo sugiere el siguiente **procedimiento para contrastar por sesgo de selección muestral**:

1. Para cada período temporal estime un modelo Probit para la ecuación:

$$\Pr(s_{it2}|z_i) = \Phi(z_{it}\delta_{t2} + \bar{z}_i\xi_{t2}), \quad t = 1, \dots, T. \quad (33)$$

Usando los resultados de la estimación **construya la inversa de los cocientes de Mills**: $\hat{\lambda}_{it2} \equiv \lambda(z_{it}\hat{\delta}_{t2} + \bar{z}_i\hat{\xi}_{t2})$.

2. Para la muestra observada, **estime (31) usando FE-2SLS**, reemplazando $E(v_{it2}|z_i, s_{it2})$ por $\hat{\lambda}_{it2}$.

Contraste por Sesgo de Selección bajo Truncamiento Incidental

3. Use un estadístico t , robusto ante la presencia de heterocedasticidad y correlación serial, para contrastar la hipótesis nula: $H_0 : \rho_1 = 0$
 - Si la hipótesis nula no se rechaza, no hay problemas de selección muestral y el estimador de FE-2SLS es consistente.
 - **Remark 1:** Note que en el paso 2. además de $\hat{\lambda}_{it2}$ se pueden agregar interacciones de la inversa del cociente de Mills con variables binarias temporales para permitir correlaciones diferentes entre los errores idiosincráticos u_{it1} y v_{it2} (i.e. un ρ_1 diferente en cada t).
 - Si se incluyen estas interacciones el contraste por sesgo de selección se realiza por un test de Wald que chequea que todos los coeficientes de las interacciones son en conjunto iguales a cero.

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- Si el contraste anterior rechaza la hipótesis nula de ausencia de sesgo de selección, entonces se requiere un procedimiento de corrección.
- El procedimiento utilizado para el contraste no funciona para corregir por sesgo de selección.
- El principal problema es la heterogeneidad no observada que aparece en el indicador de selección (c_{i2}). Esto provoca que los errores del mecanismo de selección estén serialmente correlacionados provocando que $E(v_{it2}|z_i, s_{i2})$ tenga una expresión muy complicada.
- Usando el mecanismo de Chamberlain (1980) o el de Mundlak (1978) se puede asumir linealidad en la esperanza condicional de c_{i1} y obtener una corrección de sesgo de selección válida.

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- Específicamente,

$$\begin{aligned}c_{i1} &= \bar{z}_i \xi_1 + a_{i1}, \\ E(a_{i1}|z_i) &= 0\end{aligned}\tag{34}$$

- Reemplazando (34) en la ecuación de interés tenemos,

$$y_{it1} = x_{it}\beta_1 + \bar{z}_i \xi_1 + a_{i1} + u_{it1} = x_{it}\beta_1 + \bar{z}_i \xi_1 + v_{it1}, \quad t = 1, \dots, T \tag{35}$$

con $v_{it1} \equiv a_{i1} + u_{it1}$ independiente en media de z_i .

- Ahora, introduciendo selección correlacionada con heterogeneidad no observada y con el error idiosincrático de la ecuación de interés podemos escribir el modelo tomando esperanzas condicionales en la ecuación de interés:

$$\begin{aligned}y_{it1} &= x_{it}\beta_1 + \bar{z}_i \xi_1 + E(v_{it1}|z_i, s_{it2}) + e_{it1} \\ E(e_{it1}|z_i, s_{it2}) &= 0, \quad t = 1, \dots, T\end{aligned}\tag{36}$$

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- Si conociéramos $E(v_{it1}|z_i, s_{it2})$ en la ecuación anterior, la consistencia del estimador de Pooled 2SLS estaría garantizada por el Supuesto P2SLS.1
- Recuerde que lo único que requiere el Supuesto P2SLS.1 es que e_{it1} sea independiente en media (condicionalmente) de z_{it} y el mecanismo de selección en t .
- **Supuesto P2SLS.2:** (i) las variables instrumentales z_{it} siempre se observan mientras que (y_{it1}, x_{it1}) solo se observan si $s_{it2} = 1$; (ii) el mecanismo de selección está dado por (28); (iii) c_{i1} satisface (34); (iv)
$$E(v_{it1}|z_i, v_{it2}) \equiv E(a_{i1} + u_{it1}|z_i, v_{it2}) = E(a_{i1} + u_{it1}|v_{it2}) = \gamma_{t1} v_{it2}, \quad t = 1, \dots, T$$

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- Usando (iii) y (iv) tenemos,

$$\begin{aligned} y_{it1} &= x_{it}\beta_1 + \bar{z}_i\xi_1 + \gamma_{t1}E(v_{it2}|z_i, s_{it2}) + e_{it1} \\ E(e_{it1}|z_i, s_{it2}) &= 0, \quad t = 1, \dots, T \end{aligned} \quad (37)$$

- Escribiendo el modelo para la muestra que observamos queda,

$$y_{it1}|(s_{it2} = 1) = x_{it}\beta_1 + \bar{z}_i\xi_1 + \gamma_{t1}\lambda_{it2} + e_{it1}, \quad t = 1, \dots, T \quad (38)$$

- Esto significa que podemos estimar β_1 , ξ_1 y $(\gamma_{11}, \gamma_{21}, \dots, \gamma_{T1})$ usando Pooled 2SLS una vez que reemplazamos λ_{it2} con $\hat{\lambda}_{it2}$.
- Podemos resumir la corrección del sesgo de selección en presencia de variables endógenas con el siguiente procedimiento.

Corrección del Sesgo de Selección bajo Truncamiento Incidental

1. Para cada período temporal **estime un modelo Probit para s_{it2} sobre $1, z_{it}, \bar{z}_i$, $i = 1, \dots, N$** , y obtenga la inversa de los cocientes de Mills, $\hat{\lambda}_{it2}$.
2. Para la muestra de datos observados, **estime la ecuación (38) con λ_{it2} reemplazada por $\hat{\lambda}_{it2}$ por Pooled 2SLS usando $1, z_{it}, \bar{z}_i, \hat{\lambda}_{it2}$ como instrumentos**. Note que (38) implica diferentes coeficientes para λ_{it2} en cada período temporal. Como antes, esto puede implementarse adicionando los términos de interacción apropiados en la regresión. Alternativamente, se puede estimar un modelo más restrictivo asumiendo $\gamma_{t1} = \gamma_1, \forall t$.
- Para realizar **inferencia estadística** en el modelo necesitamos estimar la matriz de varianzas y covarianzas de los estimadores del paso 2. de arriba.

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- Definamos los regresores generados y los instrumentos para el período t como: $\hat{w}_{it} = (x_{it}, \bar{z}_i, 0, \dots, 0, \hat{\lambda}_{it2}, 0, \dots, 0)$ y $\hat{h}_{it} = (z_{it}, \bar{z}_i, 0, \dots, 0, \hat{\lambda}_{it2}, 0, \dots, 0)$, respectivamente. Sin el “^” estos vectores están evaluados en los parámetros poblacionales.
- El vector de parámetros de la ecuación de interés es:
 $\theta = (\beta'_1, \xi'_1, \gamma'_{11}, \dots, \gamma'_{T1})'$.
- El vector de parámetros de la ecuación de selección es: $\pi_t = (\delta'_{t2}, \xi'_{t2})'$, y $\pi = (\pi'_1, \pi'_2, \dots, \pi'_T)'$.

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- El estimador de Pooled 2SLS es,

$$\begin{aligned}\hat{\theta} &= \left[\left(\sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{w}'_{it} \hat{h}_{it} \right) \left(\sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{h}_{it} \right)^{-1} \right. \\ &\quad \times \left. \left(\sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{w}_{it} \right) \right]^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{w}'_{it} \hat{h}_{it} \right) \\ &\quad \times \left(\sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{h}_{it} \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} y_{it1} \right) \quad (39)\end{aligned}$$

- Sustituyendo $y_{it1} = w_{it}\theta + e_{it1} = \hat{w}_{it}\theta + (w_{it} - \hat{w}_{it})\theta + e_{it1}$ en (39) tenemos,

Corrección del Sesgo de Selección bajo Truncamiento Incidental

$$\begin{aligned}\sqrt{N}(\hat{\theta} - \theta) &= \left[\left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{w}'_{it} \hat{h}_{it} \right) \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{h}_{it} \right)^{-1} \right. \\ &\times \left. \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{w}_{it} \right) \right]^{-1} \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{w}'_{it} \hat{h}_{it} \right) \\ &\times \left(N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{h}_{it} \right)^{-1} \\ &\times \left(N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} [(w_{it} - \hat{w}_{it})\theta + e_{it1}] \right) \\ &\times (C'D^{-1}C)^{-1} C'D^{-1} \\ &\times \left(N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} [(w_{it} - \hat{w}_{it})\theta + e_{it1}] \right) + o_p(1)\end{aligned}\tag{40}$$

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- En la expresión anterior $C \equiv E(\sum_{t=1}^T s_{it2} h'_{it} w_{it})$ y $D \equiv E(\sum_{t=1}^T s_{it2} h'_{it} h_{it})$.
- Usando Wooldridge (2002) y $E(e_{it1}|h_{it}, s_{it2}) = 0$,

$$\begin{aligned} & \left(N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} [(w_{it} - \hat{w}_{it})\theta + e_{it1}] \right) \\ &= N^{-1/2} \sum_{i=1}^N \left[\sum_{t=1}^T s_{it2} h'_{it} e_{it1} - F\psi_i(\pi) \right] + o_p(1) \end{aligned} \quad (41)$$

donde $F = E[\sum_{t=1}^T s_{it2} h'_{it} (\theta' \nabla_{\pi} w'_{it})]$, $\nabla_{\pi} w'_{it}$ es el Jacobiano de w'_{it} con respecto a π y $\psi_i(\pi)$ depende de la esperanza matemática de Hesianos y funciones score de la estimación del Probit del primer paso.

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- Combinando (41) con (40) tenemos,

$$\begin{aligned} & \sqrt{N}(\hat{\theta} - \theta) \\ &= (C'D^{-1}C)^{-1}C'D^{-1} \left(N^{-1/2} \sum_{i=1}^N \left[\sum_{t=1}^T s_{it2} h'_{it} e_{it1} - F\psi_i(\pi) \right] \right) \\ &+ o_p(1) \end{aligned} \quad (42)$$

- Entonces,

$$\sqrt{N}(\hat{\theta} - \theta) \xrightarrow{d} \text{Normal}[0, (C'D^{-1}C)^{-1}C'D^{-1}GD^{-1}C(C'D^{-1}C)^{-1}] \quad (43)$$

donde

$$G = \text{Var} \left(\sum_{t=1}^T s_{it2} h'_{it} e_{it1} - F\psi_i(\pi) \right) \equiv \text{Var}[g_i(\theta, \pi)].$$

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- La estimación consistente de $\text{Avar}[\sqrt{N}(\hat{\theta} - \theta)]$ se obtiene reemplazando los parámetros desconocidos por sus estimadores consistentes.
- Estimadores consistentes de C , D , y G viene dados por:

$$\hat{C} \equiv N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{w}_{it} \quad (44)$$

$$\hat{D} = N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{h}_{it} \quad (45)$$

$$\hat{G} = N^{-1} \sum_{i=1}^N \sum_{t=1}^T \hat{g}_i \hat{g}_i' \quad (46)$$

donde $\hat{g}_i = \sum_{t=1}^T s_{it2} \hat{h}'_{it} \hat{e}_{it1} - \hat{F} \hat{\psi}_i$, $\hat{e}_{it1} = y_{it1} - \hat{w}_{it} \hat{\theta}$, y $\hat{F} = N^{-1} \sum_{i=1}^N \sum_{t=1}^T s_{it2} \hat{h}'_{it} (\hat{\theta}' \nabla_{\pi} \hat{w}_{it})$.

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- En la expresión anterior para cada (i, t) ,
 $\hat{\theta}' \nabla_{\pi} \hat{w}'_{it} = (0, \dots, 0, -\hat{\gamma}_{t1} q_{it} \hat{\lambda}_{it2} (q_{it} \hat{\pi}_t + \hat{\lambda}_{it2}), 0, \dots, 0,)$ con $q_{it} \equiv (z_{it}, \bar{z}_i)$ y donde $\lambda_{it2}(q_{it}\pi_t + \lambda_{it2})$ es la derivada de la inversa del cociente de Mills.
- Con estos resultados tenemos,

$$\hat{F} = -N^{-1} \sum_{i=1}^N \sum_{t=1}^T [0, \dots, 0, s_{it2} \hat{h}'_{it} \hat{\gamma}_{t1} q_{it} \hat{\lambda}_{it2} (q_{it} \hat{\pi}_t + \hat{\lambda}_{it2}), 0, \dots, 0,] \quad (47)$$

- Desde los resultados de la estimación del Probit de la primera etapa, para cada (i, t) , tenemos,

$$\hat{\psi}_{it} = \hat{H}_t^{-1} \{ \Phi(q_{it} \hat{\pi}_t) [1 - \Phi(q_{it} \hat{\pi}_t)] \}^{-1} \phi(q_{it} \hat{\pi}_t) q'_{it} [s_{it2} - \Phi(q_{it} \hat{\pi}_t)] \quad (48)$$

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- En la expresión anterior

$$\hat{H}_t^{-1} = N^{-1} \sum_{i=1}^N \{\Phi(q_{it}\hat{\pi}_t)[1 - \Phi(q_{it}\hat{\pi}_t)]\}^{-1} [\phi(q_{it}\hat{\pi}_t)]^2 q'_{it} q_{it} \quad (49)$$

es la estimación consistente de la esperanza del Hesiano y $\hat{\pi}_t$ es estimador de máxima verosimilitud del modelo Probit de s_{it2} contra q_{it} , $i = 1, \dots, N$.

- Para cada i hay que “apilar” los $\hat{\psi}_{it}$ para obtener el $\hat{\psi}_i$ que se usa en la ecuación (46).

Truncamiento Incidental

- Suponga que y y z tienen una distribución bivariada con correlación ρ . Nosotros estamos interesados en la distribución de y dado que z excede un determinado valor. Esto es, la función de densidad conjunta de y y z es:

$$f(y, z \mid z > a) = \frac{f(y, z)}{\text{Prob}(z > a)}$$

- **Teorema 20.4 (Greene, 1997, Cap. 20, p. 975):** Si y y z tienen una distribución normal bivariada con medias μ_y y μ_z , desviaciones estándar σ_y y σ_z y correlación ρ , entonces:
- $E(y \mid z > a) = \mu_y + \rho\sigma_y\lambda(\alpha_z)$, $\text{Var}(y \mid z > a) = \sigma_y^2 [1 - \rho^2\delta(\alpha_z)]$,
- Donde: $\alpha_z = (a - \mu_z)/\sigma_z$, $\lambda(\alpha_z) = \phi(\alpha_z)/[1 - \Phi(\alpha_z)]$, $\delta(\alpha_z) = \lambda(\alpha_z)[\lambda(\alpha_z) - \alpha_z]$.

Truncamiento Incidental

- En economía el caso emblemático es el modelo de oferta de trabajo de las mujeres (Gronau, 1974; Heckman, 1976). Este modelo consiste de dos ecuaciones, una **ecuación de salarios** que representa la **diferencia entre el salario de mercado de una persona y su salario de reserva**, en función de características tales como la edad, educación, experiencia etc.
- La segunda ecuación es una **ecuación de horas deseadas de trabajo** que depende del salario, de la presencia de hijos pequeños, del estado civil, etc.
- **El problema del truncamiento es que en la segunda ecuación observamos las horas reales solo si la persona está trabajando.** Esto es, solo si el salario de mercado excede al salario de reserva. En este caso se dice que **la variable horas en la segunda ecuación está incidentalmente truncada.**

Truncamiento Incidental

- Para poner este ejemplo en un marco general de análisis, digamos que la ecuación que determina la selección muestral es:
$$z_i^* = \gamma' w_i + u_i,$$
- y la ecuación de interés es:
$$y_i = \beta' x_i + \varepsilon_i$$
- La regla es que y_i es observada solo cuando z_i^* es mayor a cero. Asumamos que u_i y ε_i tienen distribución normal bivariada con medias iguales a cero y coeficiente de correlación ρ .

Truncamiento Incidental

- Aplicando teorema 20.4:

$$(y_i \mid y_i \text{ es observada}) = E[y_i \mid z_i^* > 0] \quad (50)$$

$$= E[y_i \mid u_i > -\gamma' w_i] \quad (51)$$

$$= \beta' x_i + E[\varepsilon_i \mid u_i > -\gamma' w_i] \quad (52)$$

$$= \beta' x_i + \rho \sigma_\varepsilon \lambda_i(\alpha_u) \quad (53)$$

$$= \beta' x_i + \beta_\lambda \lambda_i(\alpha_u), \quad (54)$$

- donde $\alpha_u = -\gamma' w_i / \sigma_u$ y $\lambda_i(\alpha_u) = \phi(\gamma' w_i / \sigma_u) / \Phi(\gamma' w_i / \sigma_u)$.

Truncamiento Incidental

- Por lo tanto:

$$y_i \mid z_i^* > 0 = E[y_i \mid z_i^* > 0] + v_i = \beta' x_i + \beta_\lambda \lambda_i(\alpha_u) + v_i$$

- Está claro de este desarrollo que estimar por OLS la ecuación de horas trabajadas, solo con los datos observados, produce **estimadores inconsistentes de β por el argumento estándar de variables omitidas.**

Truncamiento Incidental

- Heckman (1979) diseñó un método de estimación en dos etapas.
- (1) Estime una ecuación probit para obtener estimaciones de γ . Para cada observación en la muestra seleccionada calcule: $\hat{\lambda}_i = \phi(\hat{\gamma}' w_i) / \Phi(\hat{\gamma}' w_i)$
- (2) Estime por OLS β y β_λ en la regresión de y sobre x y $\hat{\lambda}$.
- Contraste la hipótesis nula que el coeficiente asociado a $\hat{\lambda}$ es cero usando un estadístico t de significación individual (completamente robusto). Si acepta la hipótesis nula, no hay evidencia de sesgo de selección. Por otro lado, **si se rechaza la hipótesis nula entonces existe sesgo de selección.**
- **La estimación del segundo paso de Heckman produce estimadores insesgados y consistentes de β .**

► Go Back

Modelo Probit

- Denotemos por $\phi(\cdot)$ y por $\Phi(\cdot)$ a las funciones de densidad y de distribución acumulada de una normal estándar.
- La función de verosimilitud para el modelo Probit en cada t es:

$$L(\delta_t; X_j) = \prod_{j=1}^N [\Phi(X_j \delta_t)]^{Y_j} [1 - \Phi(X_j \delta_t)]^{1-Y_j}$$

- Donde Y_j es una variable binaria que adopta el valor 1 si j está en la muestra seleccionada.

Modelo Probit

- Y su logaritmo natural es:

$$\ell = \sum_{j=1}^N \{Y_j \log [\Phi(X_j \delta_t)] + (1 - Y_j) \log [1 - \Phi(X_j \delta_t)]\}$$

- Las condiciones de primer orden son:

$$S(\hat{\delta}_t) = \frac{\partial \ell}{\partial \hat{\delta}_t} = \sum_{j=1}^N \left(\frac{Y_j - \Phi(X_j \hat{\delta}_t)}{\Phi(X_j \hat{\delta}_t) [1 - \Phi(X_j \hat{\delta}_t)]} \phi(X_j \hat{\delta}_t) \right) X_j = 0$$

Modelo Probit

- La matriz de información de Fisher es:

$$I(\hat{\delta}_t) = E \left(-\frac{\partial^2 \ell}{\partial \hat{\delta} \partial \hat{\delta}_t} \right) = \sum_{j=1}^N \left(\frac{\phi(X_j \hat{\delta}_t)^2}{\Phi(X_j \hat{\delta}_t) [1 - \Phi(X_j \hat{\delta}_t)]} \right) X_j X_j'$$

- Por lo tanto: $\hat{r}_{jt} = \left[I(\hat{\delta}_t) \right]^{-1} S(\hat{\delta}_t)$, $t = 1, 2, \dots, T$.
- Donde $\hat{\delta}_t$ es de dimensión $(1 + TK) \times 1$; $\left[I(\hat{\delta}_t) \right]^{-1}$ es de dimensión $(1 + TK) \times (1 + TK)$; y $S(\hat{\delta}_t)$ es de dimensión $(1 + TK) \times 1$.
- Construya el vector \hat{r}_j de dimensión $T(1 + TK) \times 1$ “stacking” $\{\hat{r}_{j1}, \hat{r}_{j2}, \dots, \hat{r}_{jT}\}$.

Modelo Tobit

- El modelo Tobit se define como sigue: $y_i = \beta'x_i + u_i$ si $\beta'x_i + u_i > 0$; e $y_i = 0$ en cualquier otro caso.
- β es un vector $k \times 1$ de parámetros, x_i es un vector $k \times 1$ de variables explicativas; y u_i son los errores que son independientes y se distribuyen con distribución normal con media cero y varianza σ^2 .
- Supongamos que tenemos N_0 observaciones para las que $y_i = 0$, y N_1 observaciones para las que $y_i > 0$. Sin pérdida de generalidad asumamos que las N_1 observaciones no cero de y_i están primero en la muestra.
- Por conveniencia definamos:

$$F_i = F(\beta'x_i, \sigma^2) = \int_{-\infty}^{\beta'x_i} \frac{1}{\sigma(2\pi)^{1/2}} e^{-t^2/2\sigma^2} dt$$

$$\text{y } f_i = f(\beta'x_i, \sigma^2) = \frac{1}{\sigma(2\pi)^{1/2}} e^{-t^2/2\sigma^2}$$

Modelo Tobit

- y_i y $\Phi_i = F_i = \int_{-\infty}^{\beta'x_i/\sigma} \frac{1}{(2\pi)^{1/2}} e^{-t^2/2} dt$; $\phi_i = \sigma f_i = \frac{1}{(2\pi)^{1/2}} e^{-(\beta'x_i)^2/2\sigma^2}$
- Son las funciones acumulada y de densidad de la normal estándar evaluadas en $\beta'x_i/\sigma$.
- Además, $\gamma_i = \frac{\phi_i}{1-\Phi_i}$,
 $Y'_1 = (y_1, y_2, \dots, y_{N_1})$,
 $X'_1 = (x_1, x_2, \dots, x_{N_1})$,
 $X'_0 = (x_{N_1+1}, x_{N_1+2}, \dots, x_N)$ y
 $\gamma'_0 = (\gamma_{N_1+1}, \gamma_{N_1+2}, \dots, \gamma_N)$.
- Para las observaciones $y_i = 0$, lo único que sabemos es:
 $Prob(y_i = 0) = Prob(u_i < -\beta'x_i) = (1 - F_i)$
- Para las observaciones $y_i > 0$, tenemos
 $Prob(y_i > 0)f_i(y_i | y_i > 0) = F_i \frac{f(y_i - \beta'x_i, \sigma^2)}{F_i} = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(1/2\sigma^2)(y_i - \beta'x_i)^2}$

- La función de verosimilitud es entonces,

$$L = \prod_{\forall y_i=0} (1 - F_i) \prod_{\forall y_i>0} \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(1/2\sigma^2)(y_i - \beta'x_i)^2}$$

- Tomando logaritmos $\log L = \sum_{\forall y_i=0} \log(1 - F_i) + \sum_{\forall y_i>0} \log \left(\frac{1}{(2\pi\sigma^2)^{1/2}} \right) - \sum_{\forall y_i>0} \frac{1}{2\sigma^2} (y_i - \beta'x_i)^2$

- Usando las siguientes expresiones:

- $\frac{\partial F_i}{\partial \beta} = f_i x_i$

- $\frac{\partial F_i}{\partial \sigma^2} = -\frac{1}{2\sigma^2} \beta' x_i f_i$

- $\frac{\partial f_i}{\partial \beta} = -\frac{1}{\sigma^2} \beta' x_i f_i x_i$

- $\frac{\partial f_i}{\partial \sigma^2} = \frac{(\beta' x_i)^2 - \sigma^2}{2\sigma^4} f_i$

- Las condiciones de primer orden quedan:

$$\frac{\partial \log L}{\partial \beta} = - \sum_{\forall y_i=0} \frac{f_i x_i}{1-F_i} + \frac{1}{\sigma^2} \sum_{\forall y_i>0} (y_i - \beta' x_i) x_i = 0 \text{ y}$$

$$\frac{\partial \log L}{\partial \sigma^2} = \frac{1}{2\sigma^2} \sum_{\forall y_i=0} \frac{\beta' x_i f_i}{1-F_i} - \frac{N_1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{\forall y_i>0} (y_i - \beta' x_i)^2 = 0$$

- Premultiplicando la primera ecuación por $\beta'/2\sigma^2$ y sumando el resultado a la segunda ecuación obtenemos:
- $\sigma^2 = \frac{1}{N_1} \sum_{\forall y_i>0} (y_i - \beta' x_i) y_i = \frac{Y_1(Y_1 - X_1\beta)}{N_1}$
- Después, multiplicando la primera ecuación por σ y despejando β :

$$\beta = (X_1' X_1)^{-1} X_1' Y_1 - \sigma (X_1' X_1)^{-1} X_0' \gamma_0$$

Modelo Tobit

- Las condiciones de segundo orden son:

- $$\frac{\partial^2 \log L}{\partial \beta \partial \beta'} = - \sum_{\forall y_i=0} \frac{f_i}{(1-F_i)^2} \left[f_i - \frac{1}{\sigma^2} (1-F_i) \beta' x_i \right] x_i x_i' - \frac{1}{\sigma^2} \sum_{\forall y_i>0} x_i x_i'$$



$$\begin{aligned} \frac{\partial^2 \log L}{\partial \sigma^2 \partial \beta} = & -\frac{1}{2\sigma^2} \sum_{\forall y_i=0} \frac{f_i}{(1-F_i)^2} \left[\frac{1}{\sigma^2} (1-F_i) (\beta' x_i)^2 - (1-F_i) \right. \\ & \left. - \beta' x_i f_i \right] x_i - \frac{1}{\sigma^4} \sum_{\forall y_i>0} (y_i - \beta' x_i) x_i \end{aligned}$$

- $$\begin{aligned} \frac{\partial^2 \log L}{\partial (\sigma^2)^2} = & \frac{1}{4\sigma^4} \sum_{\forall y_i=0} \frac{f_i}{(1-F_i)^2} \left[\frac{1}{\sigma^2} (1-F_i) (\beta' x_i)^3 - 3(1-F_i) \beta' x_i - (\beta' x_i)^2 f_i \right] \\ & + \frac{N_1}{2\sigma^4} - \frac{1}{\sigma^6} \sum_{\forall y_i>0} (y_i - \beta' x_i)^2 \end{aligned}$$

Modelo Tobit

- Defina la score function como $S(\hat{\beta}) = \frac{\partial \log L}{\partial \hat{\beta}}$
- Y la matriz de información de Fisher como:

$$I(\hat{\beta}) = E \left(-\frac{\partial^2 \log L}{\partial \hat{\beta} \partial \hat{\beta}'} \right)$$

- Por lo tanto

$$\hat{r}_{jt} = \left[I(\hat{\beta}) \right]^{-1} S(\hat{\beta})$$

► Go Back

Modelo Probit Bivariante

- En economía la motivación para utilizar este tipo de modelos proviene de la racionalización de que los resultados binarios observados pueden no reflejar la elección de un solo agente económico sino de la decisión conjunta de dos.
- Por ejemplo, Gunderson (1974) discute modelos estadísticos para estimar la probabilidad de que un empleado que recibe entrenamiento en el trabajo sea retenido por la empresa después de dicho entrenamiento.

Modelo Probit Bivariante

- En esta situación, el empleador debe decidir si hacer o no una oferta de trabajo, y el empleado debe decidir si busca o no que le hagan la oferta.
- Las decisiones individuales no son observables y lo único que uno observa es si el empleado continúa trabajando después de completar su entrenamiento o no.
- En particular, las decisiones individuales se modelan como probits univariantes.
- Mientras que las dos decisiones tomadas conjuntamente se modelan como un probit bivariante.
- Por ejemplo, considere las siguientes decisiones individuales:
 - $y_1^* = x\beta_1 + v_1, y_i = 1 \iff y_1^* > 0$
 - $y_2^* = x\beta_2 + v_2, y_2 = 1 \iff y_2^* > 0$

Modelo Probit Bivariante

- Consideremos el caso de Gunderson, donde solo observamos si o no, y_1 e y_2 son iguales a uno. Definamos $z_j = y_{1j}y_{2j}$ para $j = 1, 2, \dots, N$.
- Como $z_j = 1$ sí y solo sí $y_{1j} = 1$ e $y_{2j} = 1$, la distribución de z_j es,
- $p_j = Pr(z_j = 1) = Pr(y_{1j} = 1, y_{2j} = 1) = F(x_j\beta_1, x_j\beta_2; \rho)$,
- $1 - p_j = Pr(z_j = 0) = Pr(y_{1j} = 0 \text{ ó } y_{2j} = 0) = 1 - F(x_j\beta_1, x_j\beta_2; \rho)$
- Donde ρ es el coeficiente de correlación entre v_1 y v_2 , y $F(\cdot)$ denota la distribución normal estándar bivariada.

Modelo Probit Bivariante

- El logaritmo de la función de verosimilitud es:

$$\begin{aligned} L(\beta_1, \beta_2, \rho) = & \sum_{j=1}^N z_j \log [F(x_j \beta_1, x_j \beta_2; \rho)] \\ & + (1 - z_j) \log [1 - F(x_j \beta_1, x_j \beta_2; \rho)] \end{aligned} \quad (55)$$

- Los estimadores de MV se obtienen maximizando esta función con respecto a:

$$\hat{\theta} = \left(\hat{\beta}'_1, \hat{\beta}'_2, \hat{\rho} \right)'$$

Modelo Probit Bivariante

- Estos modelos tienen importantes implicancias para los problemas de sesgo de selección muestral.
- Supongamos que al modelo de Gunderson le agregamos una ecuación salarial:
 $y_3 = x\beta_3 + v_3,$
- Donde y_3 denota salarios. Supongamos que $v = [v_1, v_2, v_3]$ tiene distribución normal trivariante con media cero y matriz de varianzas covarianzas igual a:

Modelo Probit Bivariante

$$\bullet \begin{pmatrix} 1 & \rho & \sigma_{13} \\ \rho & 1 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{pmatrix}$$

- Dado que observamos solo los salarios de los empleados que siguen trabajando en la empresa después de terminar el entrenamiento, la esperanza condicional del error de la ecuación de salarios es:

Modelo Probit Bivariante



$$\begin{aligned} E(v_3 \mid y_1^* > 0, y_2^* > 0) &= E(v_3 \mid v_1 > -x\beta_1, v_2 > -x\beta_2) \\ &= \sigma_{23} \left[\frac{\phi(x\beta_1) \Phi \left[x(\beta_2 - \rho\beta_1) / (1 - \rho^2)^{1/2} \right]}{F(x\beta_1, x\beta_2; \rho)} \right] \\ &+ \sigma_{23} \left[\frac{\phi(x\beta_2) \Phi \left[x(\beta_1 - \rho\beta_2) / (1 - \rho^2)^{1/2} \right]}{F(x\beta_1, x\beta_2; \rho)} \right] \\ &= \sigma_{23}\lambda_1 + \sigma_{23}\lambda_2 \end{aligned} \quad (56)$$

- Por lo tanto para corregir la ecuación de salarios por sesgo de selección muestral hay que agregarle estos dos términos de la esperanza condicional.

Modelo Probit Bivariante

- Lo importante de este procedimiento es que si z_j hubiera sido modelada como un modelo probit univariante el (incorrecto) término de corrección por sesgo de selección muestral sería la inversa del cociente de Mills.
- Como el proceso de decisión involucra un modelo probit bivariado, aparecen dos términos para corregir por el sesgo de selección muestral.

Modelo Probit Bivariante

- En la práctica este procedimiento sería:
- (1) Estimar el modelo probit bivariante maximizando (55) y obtener los estimadores de β_1, β_2 , y ρ ;
- (2) Construir los términos expresados entre corchetes en (56) (i.e. λ_1 y λ_2); y
- (3) Estimar por OLS: $y_3 = x\beta_3 + \ell_1\lambda_1 + \ell_2\lambda_2 + v_3$

Modelo Probit Bivariante

- Para más detalles puede consultar
- Poirier, D (1980) “Partial observability in bivariate probit models”, Journal of Econometrics, 12, pp. 209-217.

► Go Back

- Recuerde (9) con: $y_{jt} - y_{js} = (x_{jt}^1 - x_{js}^1) \beta + (u_{jt} - u_{js})$, y $s_{jt} = 1 [X_j \delta_t + v_{jt} \geq 0]$ para $\tau = t, s$.
- Entonces,
$$E(y_{it} - y_{js} \mid X_j, v_{jt} \geq -H_{jt}, v_{js} \geq -H_{js}) = (x_{jt}^1 - x_{js}^1) \beta + \ell_{ts} \lambda_{jts} + \ell_{st} \lambda_{jst},$$

- Donde,

$$\lambda_{jts} = \frac{\phi[X_j\delta_t] \Phi\left[\frac{X_j\delta_s - \rho_{ts}X_j\delta_t}{(1-\rho_{ts}^2)^{1/2}}\right]}{\Phi_2[X_j\delta_t, X_j\delta_s, \rho_{ts}]}$$

$$\lambda_{jst} = \frac{\phi[X_j\delta_s] \Phi\left[\frac{X_j\delta_t - \rho_{ts}X_j\delta_s}{(1-\rho_{ts}^2)^{1/2}}\right]}{\Phi_2[X_j\delta_t, X_j\delta_s, \rho_{ts}]}$$

La Matriz de Varianzas

- La estimación en dos pasos es como sigue. Primero estime $\bar{\omega}_{ts} = (\delta'_t, \delta'_s, \rho_{ts})'$ con $\hat{\bar{\omega}}_{ts} = (\hat{\delta}'_t, \hat{\delta}'_s, \hat{\rho}_{ts})'$
- Usando un modelo Probit bivariado con las observaciones de (s_{jt}, s_{js}, X_j) .
- Segundo, para la submuestra con $s_{jt} = s_{js} = 1$, estime por POLS $\Delta y_{jts} = (y_{jt} - y_{js})$ sobre $\Delta x^1_{jts} = (x^1_{jt} - x^1_{js})$ y $(\hat{\lambda}_{jts}, \hat{\lambda}_{jst})$.

La Matriz de Varianzas

- El segundo paso da estimaciones consistentes de los parámetros de interés β y de los coeficientes que acompañan a los términos de corrección del sesgo de selección muestral ℓ_{ts} y ℓ_{st} .
- Definamos $R_{jts} \equiv (\Delta x_{jts}^1, \lambda_{jts}, \lambda_{jst})'$ y las condiciones de primer orden en el segundo paso son:
- $$\frac{1}{N} \sum_j s_{jt} s_{js} \left\{ \Delta y_{jts} - \Delta x_{jts}^1 \hat{\beta} - \hat{\ell}_{ts} \hat{\lambda}_{jts} - \hat{\ell}_{st} \hat{\lambda}_{jst} \right\} R_{jts} = 0$$

La Matriz de Varianzas

- Defina las siguientes equivalencias:
- $\hat{\pi}_{ts} \equiv \left(\hat{\beta}, \hat{\ell}_{ts}, \hat{\ell}_{st} \right)'$
- $\pi_{ts} \equiv \left(\beta, \ell_{ts}, \ell_{st} \right)'$
- $e_{jts} \equiv \Delta y_{jts} - \Delta x_{jts}^1 \beta - \ell_{ts} \lambda_{jts} - \ell_{st} \lambda_{jst}$
- Con estas definiciones se puede mostrar que:

$$\begin{aligned} \sqrt{N} \left(\widehat{\omega}_{ts} - \bar{\omega}_{ts} \right) &\xrightarrow{p} \frac{1}{\sqrt{N}} \sum_j I_{\bar{\omega}_{ts}}^{-1} \begin{pmatrix} X_j' (q_{jt} \phi_{jt} \Phi_{jts} / \Phi_{2,jts}) \\ X_j' (q_{js} \phi_{st} \Phi_{jts} / \Phi_{2,jts}) \\ q_{jt} q_{js} \phi_{2,jts} / \Phi_{2,jts} \end{pmatrix} \\ &\equiv \frac{1}{\sqrt{N}} \sum_j \Lambda_j \end{aligned}$$

La Matriz de Varianzas

- Donde $I_{\bar{\omega}_{TS}}^{-1}$ es la inversa de la matriz de información de Fisher del modelo Probit bivariado para $\bar{\omega}_{ts}$,
- $\phi_{jt} \equiv \phi[q_{jt}X_j\delta_t]$, $\phi_{js} \equiv \phi[q_{js}X_j\delta_s]$
- $\Phi_{jts} \equiv \Phi \left[(q_{js}X_j\delta_s - \rho_{jts}^* q_{jt}X_j\delta_t) / (1 - \rho_{jts}^{*2})^{1/2} \right]$
- $\Phi_{jts} \equiv \Phi \left[(q_{jt}X_j\delta_t - \rho_{jts}^* q_{js}X_j\delta_s) / (1 - \rho_{jts}^{*2})^{1/2} \right]$
- $\Phi_{2,jts} \equiv \Phi [q_{jt}X_j\delta_t, q_{js}X_j\delta_s, \rho_{jts}^*]$ y $\phi_{2,jts} \equiv \phi [q_{jt}X_j\delta_t, q_{js}X_j\delta_s, \rho_{jts}^*]$,
 $q_{jt} \equiv 2s_{jt} - 1$, $q_{js} \equiv 2s_{js} - 1$, $\rho_{jts}^* \equiv q_{jt}q_{js}\rho_{ts}$.

La Matriz de Varianzas

- Además,

$$\sqrt{N}(\hat{\pi}_{ts} - \pi_{ts}) \xrightarrow{P} E(s_t s_s R_{ts} R'_{ts})^{-1} \frac{1}{\sqrt{N}} \sum_j \{s_{jt} s_{js} e_{jts} R_{jts} + A \Lambda_j\}$$

- donde,

$$A \equiv E \left\{ s_t s_s \left[\left(-\ell_{ts} \frac{\partial \lambda_{ts}}{\partial X \delta_t} - \ell_{ts} \frac{\partial \lambda_{st}}{\partial X \delta_t} \right) R_{ts} X \right. \right. \\ \left. \left. \left(-\ell_{ts} \frac{\partial \lambda_{ts}}{\partial X \delta_s} - \ell_{ts} \frac{\partial \lambda_{st}}{\partial X \delta_s} \right) R_{ts} X \left(-\ell_{ts} \frac{\partial \lambda_{ts}}{\partial X \rho_{ts}} - \ell_{ts} \frac{\partial \lambda_{st}}{\partial X \rho_{ts}} \right) R_{ts} \right] \right\}$$

La Matriz de Varianzas

- Por lo tanto, $\sqrt{N}(\hat{\pi}_{ts} - \pi_{ts}) \rightarrow N(0, \Gamma)$
- Donde

$$\begin{aligned}\Gamma &= E(s_t s_s R_{ts} R'_{ts})^{-1} \times E\{(s_t s_s e_{ts} R_{ts} + A\Lambda)(s_t s_s e_{ts} R_{ts} + A\Lambda)'\} \\ &\quad \times E(s_t s_s R_{ts} R'_{ts})^{-1}\end{aligned}$$

- El término $A\Lambda$ es el efecto del primer paso en el segundo.

La Matriz de Varianzas

- La estimación de la inversa de la matriz de información de Fisher puede obtenerse usando el estimador de Berndt et al.(1974)

$$\hat{I}_{\varpi_{ts}}^{-1} = \left\{ \frac{1}{N} \sum_j \begin{bmatrix} X_j' (q_{jt} \phi_{jt} \Phi_{jts} / \Phi_{2,jts}) \\ X_j' (q_{js} \phi_{st} \Phi_{jts} / \Phi_{2,jts}) \\ q_{jt} q_{js} \phi_{2,jts} / \Phi_{2,jts} \end{bmatrix} \begin{bmatrix} X_j' (q_{jt} \phi_{jt} \Phi_{jts} / \Phi_{2,jts}) \\ X_j' (q_{js} \phi_{st} \Phi_{jts} / \Phi_{2,jts}) \\ q_{jt} q_{js} \phi_{2,jts} / \Phi_{2,jts} \end{bmatrix}' \right\}^{-1}$$

► Go Back

Sample Selection: Type III Tobit Model

- Consider the case where the selection equation is of the censored Tobit form.
- The population model is

$$y_1 = x_1\beta_1 + u_1 \quad (57)$$

$$y_2 = \max(0; x\delta_2 + v_2) \quad (58)$$

where (x, y_2) is always observed in the population but y_1 is observed only when $y_2 > 0$.

- A standard example occurs when y_1 is the log of the hourly wage offer and y_2 is weekly hours of labor supply.

Sample Selection: Type III Tobit Model

- **Assumption:** (a) (x, y_2) is always observed in the population, but y_1 is observed only when $y_2 > 0$; (b) $(u_1; v_2)$ is independent of x ; (c) $v_2 \sim \text{Normal}(0; \tau_2^2)$; and (d) $E(u_1|v_2) = \gamma_1 v_2$.

Sample Selection: Type III Tobit Model

- Define the selection indicator as $s_2 = 1$ if $y_2 > 0$, and $s_2 = 0$ otherwise.
- Since s_2 is a function of x and v_2 , it follows immediately that

$$E(y_1|x; v_2; s_2 = 1) = x_1\beta_1 + \gamma_1 v_2 \quad (59)$$

- This equation means that, if we could observe v_2 , then an OLS regression of y_1 on x_1 , and v_2 using the selected subsample would consistently estimate $(\beta_1; \gamma_1)$.
- v_2 cannot be observed when $y_2 = 0$ (because when $y_2 = 0$, we only know that $v_2 \leq x\delta_2$, for $y_2 > 0$, $v_2 = y_2 - x\delta_2$).
- If we knew δ_2 , we would know v_2 whenever $y_2 > 0$.
- It seems reasonable that, because δ_2 can be consistently estimated by Tobit on the whole sample, we can replace v_2 with consistent estimates.

Sample Selection: Type III Tobit Model

- **Estimation Procedure:** (a) Estimate equation (58) by standard Tobit using all N observations. For $y_{i2} > 0$ (say $i = 1, 2, \dots, N_1$), define

$$\hat{v}_{i2} = y_{i2} - x_i \hat{\delta}_2 \quad (60)$$

- (b) Using observations for which $y_{i2} > 0$, estimate $(\beta_1; \gamma_1)$ by the OLS regression: y_{i1} on x_{i1} , and \hat{v}_{i2} $i = 1, 2, \dots, N_1$
- This regression produces consistent and \sqrt{N} asymptotically normal estimators of $(\beta_1; \gamma_1)$.

► Go Back

Econometría de Datos de Panel

Maestrías en Economía y Econometría

Lecture 4

Agenda

- 1 Modelos de Respuesta Binaria
 - Modelo Probit de Efectos no Observables
 - Modelo Logit de Efectos no Observables

- 2 Modelos Dinámicos de Respuesta Binaria con Efectos no Observables

Modelos de Respuesta Binaria

- En el caso de paneles, utilizar un modelo de probabilidad lineal cuando la variable dependiente es binaria tiene los mismos problemas que en el caso de una muestra de corte transversal. [▶ Go](#)
- Por lo tanto comenzaremos discutiendo directamente los modelos Probit y Logit.
- Para ilustrar estos modelos veremos el caso más sencillo: Pooled Probit y Logit.
- Supongamos que el modelo es: $P(y_{jt} = 1 \mid x_{jt}) = G(x_{jt}\beta)$ $t = 1, 2, \dots, T$.
- Donde $G(\cdot)$ es una función conocida que toma valores en el intervalo abierto cero-uno. x_{jt} son las variables explicativas del modelo que pueden incluir dummies temporales, interacciones de estas variables con variables constantes o no en el tiempo etc.

Modelos de Respuesta Binaria

- Dada esta especificación uno puede obtener estimaciones consistentes de β maximizando el logaritmo de la función de verosimilitud:

$$\sum_{j=1}^N \sum_{t=1}^T \{y_{jt} \log G(x_{jt}\beta) + (1 - y_{jt}) \log [1 - G(x_{jt}\beta)]\}$$

- Sin supuestos adicionales se debe utilizar un estimador robusto de la matriz de varianzas y covarianzas.

Modelos de Respuesta Binaria

- En particular un estimador simple de esta matriz viene dado por:

$$\left[\sum_{j=1}^N \sum_{t=1}^T A_{jt}(\hat{\beta}) \right]^{-1} \left[\sum_{j=1}^N S_j(\hat{\beta}) S_j(\hat{\beta})' \right] \left[\sum_{j=1}^N \sum_{t=1}^T A_{jt}(\hat{\beta}) \right]^{-1} \quad (1)$$

- Donde,

$$A_{jt}(\hat{\beta}) = \frac{\left\{ g(x_{jt}\hat{\beta}) \right\}^2 x_{jt}' x_{jt}}{G(x_{jt}\hat{\beta}) \left[1 - G(x_{jt}\hat{\beta}) \right]}$$

Modelos de Respuesta Binaria

- y por otro lado se tiene que:

$$S_j(\hat{\beta}) = \sum_{t=1}^T S_{jt}(\hat{\beta}) = \sum_{t=1}^T \frac{g(x_{jt}\hat{\beta}) x'_{jt} [y_{jt} - G(x_{jt}\hat{\beta})]}{G(x_{jt}\hat{\beta}) [1 - G(x_{jt}\hat{\beta})]}$$

con

$$g(x_{jt}\hat{\beta}) = \frac{\partial G(x_{jt}\hat{\beta})}{\partial x_{jt}\hat{\beta}}$$

Modelos de Respuesta Binaria

- Note que $S_{jt}(\hat{\beta})$ es la denominada score function (i.e. la derivada primera del logaritmo de la función de verosimilitud); mientras que $A_{jt}(\hat{\beta})$ es igual a menos la matriz Hesiana promedio estimada (i.e. la matriz de información de Fisher).
- El estimador en (1) contiene los productos cruzados de la función score y por lo tanto es robusta ante la presencia de heterocedasticidad y correlación serial.
- Si en todas las expresiones anteriores reemplazamos a $G(\cdot)$ y $g(\cdot)$ por las funciones $\Phi(\cdot)$ y $\phi(\cdot)$ respectivamente, tenemos **el modelo Probit**. $\Phi(\cdot)$ y $\phi(\cdot)$ son las funciones de distribución acumulada y de densidad de la normal estándar, respectivamente.
- Si en todas las expresiones anteriores reemplazamos a $G(\cdot)$ y $g(\cdot)$ por las funciones $\Lambda(\cdot)$ y $\lambda(\cdot)$ respectivamente, tenemos **el modelo Logit**. $\Lambda(\cdot)$ y $\lambda(\cdot)$ son las funciones de distribución acumulada y de densidad de la logística, respectivamente.

- 1 Modelos de Respuesta Binaria
 - Modelo Probit de Efectos no Observables
 - Modelo Logit de Efectos no Observables
- 2 Modelos Dinámicos de Respuesta Binaria con Efectos no Observables

Modelo Probit de Efectos no Observables

- Este modelo tiene como supuesto principal que:

$$P(y_{jt} = 1 \mid X_j, c_j) = P(y_{jt} = 1 \mid x_{jt}, c_j) = \Phi(x_{jt}\beta + c_j), t = 1, 2, \dots, T \quad (2)$$

- La primera igualdad dice que x_{jt} es estrictamente exógeno condicional en c_j . Esto deja afuera a los modelos dinámicos vistos en el curso.
- La segunda igualdad de la condición anterior es el supuesto estándar del Probit.
- Además del supuesto de exogeneidad estricta, necesitamos el supuesto estándar de que la variable dependiente es independiente condicional en (X_j, c_j) . Esto es:

$$y_{j1}, y_{j2}, \dots, y_{jT} \text{ son independientes condicionando en } (X_j, c_j). \quad (3)$$

Modelo Probit de Efectos no Observables

- Bajo estos dos supuestos podemos construir la función de verosimilitud como:

$$L(c_j, \beta) = \prod_{j=1}^N \prod_{t=1}^T \Phi(x_{jt}\beta + c_j)^{y_{jt}} [1 - \Phi(x_{jt}\beta + c_j)]^{1-y_{jt}}$$

- Idealmente, uno podría estimar los parámetros del modelo sin restringir la relación entre c_j y x_{jt} . Con este espíritu, un modelo Probit de efectos fijos trata a los efectos no observables c_j como parámetros a estimar.
- El logaritmo de la función de verosimilitud es simplemente el logaritmo de la expresión anterior.
- Desafortunadamente, además de ser una expresión complicada de estimar, en este modelo estimar juntos los β y los c_j lleva a obtener estimadores inconsistentes para β con T fijo y $N \rightarrow \infty$.
- Para poder obtener estimaciones consistentes de β necesitamos hacer algún supuesto acerca de la relación entre c_j y x_{jt} .

Modelo Probit de Efectos no Observables

- El modelo tradicional adiciona el siguiente supuesto:

$$c_j|X_j \sim \text{Normal}(0, \sigma_c^2) \quad (4)$$

- Este es un supuesto fuerte porque implica que c_j y X_j son independientes y que c_j tiene distribución normal.
- Bajo los supuestos (2), (3) y (4) existe un enfoque de máxima verosimilitud condicional para estimar β y σ_c^2 .
- Como los c_j no se observan, no pueden aparecer en la función de verosimilitud. En este caso, construimos la función de verosimilitud de $(y_{j1}, y_{j2}, \dots, y_{jT})$ condicional a X_j , lo que requiere eliminar los c_j vía integrarlos entre menos y más infinito (i.e. tienen distribución normal). Por lo tanto: $L(c_j, \beta) =$

$$\int_{-\infty}^{\infty} \left[\prod_{j=1}^N \prod_{t=1}^T \Phi(x_{jt}\beta + c_j)^{y_{jt}} [1 - \Phi(x_{jt}\beta + c_j)^{1-y_{jt}}] \right] (1/\sigma_c) \phi(c/\sigma_c) dc$$

Modelo Probit de Efectos no Observables

- Tomando el logaritmo de la expresión anterior obtenemos la función objetivo a maximizar para obtener estimadores de β y σ_c^2 consistentes y asintóticamente normales.
- La estimación por máxima verosimilitud condicional recibe el nombre en la literatura de modelo Probit de efectos aleatorios.
- Los supuestos (3) y (4) son bastante fuertes y es posible relajarlos. [▶ Go](#)
Consideremos relajar el supuesto (3).
- Bajo los supuestos (2) y (4) solamente,
$$P(y_{jt} = 1 \mid X_j) = P(y_{jt} = 1 \mid x_{jt}) = \Phi(x_{jt}\beta_c), \quad t = 1, 2, \dots, T.$$
- Donde $\beta_c = \beta / (1 + \sigma_c^2)^{1/2}$. [▶ Go](#)

Modelo Probit de Efectos no Observables

- Acabamos de mostrar que Pooled Probit de y sobre X estima consistentemente: $\beta_c = \beta / (1 + \sigma_c^2)^{1/2}$ en lugar de β , por lo tanto, el problema de los efectos no observables es más grave en el modelo Probit que en modelos lineales ya que aun cuando c_j y X_j sean independientes los coeficientes estimados son inconsistentes.
- Sin embargo, usualmente uno está interesado en el efecto parcial de las variables independientes sobre la variable dependiente y en estos casos estimar β_c es igual de bueno que estimar β .
- Para ilustrar este punto calculemos el efecto parcial para una variable x_{jt} continua,

$$\partial P(y_{jt} = 1 \mid X_j, c_j) / \partial x_{jt} = \beta_j \phi(x_{jt} \beta + c_j)$$

Modelo Probit de Efectos no Observables

- Usualmente lo que uno hace en la práctica es calcular el efecto parcial promedio. Esto es, se reemplazan las variables x_{jt} por sus promedios y se promedia la ecuación anterior a lo largo de la distribución de c en la población. Es decir: $E_c [\beta_j \phi (\bar{x}\beta + c_j)] = \left[\beta_j / (1 + \sigma_c^2)^{1/2} \right] \phi \left[\bar{x}\beta / (1 + \sigma_c^2)^{1/2} \right]$
- En otras palabras, pooled Probit de y sobre X estima consistentemente los efectos parciales promedio, que es lo que uno usualmente quiere.
- Sin embargo, como sucedía en el caso lineal, muchas veces el punto de introducir los efectos no observables es permitir explícitamente que estén correlacionados con las variables explicativas.
- Chamberlain (1980) permite esta correlación asumiendo una distribución normal condicional con esperanza lineal y varianza constante.

Modelo Probit de Efectos no Observables

- La versión de Mundlak (1978) del supuesto de Chamberlain es:

$$c_j | X_j \sim \text{Normal}(\psi + \bar{X}_j \xi, \sigma_a^2) \quad (5)$$

- Donde \bar{X}_j es el promedio de las x_{jt} , $t = 1, 2, \dots, T$ y σ_a^2 es la varianza de a_j en la ecuación: $c_j = \psi + \bar{X}_j \xi + a_j$
- Chamberlain permite una mayor generalidad reemplazando \bar{X}_j con X_j en el supuesto anterior. Esta mayor generalización tiene como contrapartida menos grados de libertad.
- Chamberlain llamó al modelo dado por (2) y (5) un modelo Probit de efectos aleatorios, y la literatura se refiere a este modelo como el modelo Probit de efectos aleatorios de Chamberlain.

Modelo Probit de Efectos no Observables

- Si los supuestos (2), (3) y (5) se cumplen, la estimación de β , ψ , ξ , y σ_a^2 es directa porque podemos escribir el modelo en forma de variable latente como:

$$y_{jt}^* = \psi + x_{jt}\beta + \bar{X}_j\xi + a_j + u_{jt}$$

- En otras palabras, adicionando \bar{X}_j a la ecuación para cada período temporal llegamos al modelo tradicional Probit con efectos aleatorios.
- Note que el efecto de asumir que la esperanza de los efectos no observables es lineal provoca el mismo efecto que en la corrección de Wooldridge en el caso de los paneles no balanceados. Allí aparecía en la ecuación de interés el término $\bar{X}_j\pi$ y ahora aparece $\bar{X}_j\xi$ debido al supuesto de Mundlak.

Modelo Probit de Efectos no Observables

- Dadas las estimaciones de ψ y ξ podemos estimar $E(c_j) = \psi + E(\bar{X}_j) \xi$ con $\hat{\psi} + \bar{X}\xi$
- Donde \bar{X} es la media muestral de \bar{X}_j .
- Por lo tanto para cualquier vector X_t , podemos estimar la probabilidad de respuesta en $E(c_j)$ como: $\Phi(\hat{\psi} + X_t\hat{\beta} + \bar{X}\hat{\xi})$
- Tomando derivadas con respecto a los elementos de X_t podemos estimar los efectos parciales de las variables sobre las probabilidades de respuesta.

Modelo Probit de Efectos no Observables

- Si solo asumimos (2) y (5):

$$P(y_{jt} = 1 \mid X_j) = \Phi \left[(\psi + x_{jt}\beta + \bar{X}_j\xi) (1 + \sigma_a^2)^{-1/2} \right] \equiv \Phi (\psi_a + x_{jt}\beta_a + \bar{X}_j\xi)$$

- Donde el subíndice a indica que el vector de parámetros ha sido multiplicado por $(1 + \sigma_c^2)^{-1/2}$.
- Se sigue inmediatamente que β_a , ψ_a , y ξ_a pueden estimarse consistentemente usando pooled Probit en y_{jt} sobre 1 , x_{jt} , y \bar{X}_j .

Modelo Probit de Efectos no Observables

- Como los y_{jt} son dependientes (no estamos asumiendo (3)) necesitamos varianzas y covarianzas robustas.
- Una vez que estimamos β_a , ψ_a , y ξ_a pueden estimarse los efectos parciales promedios usando el promedio a través de j de: $\hat{\beta}_{aj}\phi\left(\hat{\psi}_a + x^0\hat{\beta}_a + \hat{X}_j\hat{\xi}_a\right)$
- Donde x^0 es un vector no aleatorio de números que se eligen como valores interesantes de las variables explicativas.

Agenda

- 1 Modelos de Respuesta Binaria
 - Modelo Probit de Efectos no Observables
 - Modelo Logit de Efectos no Observables

- 2 Modelos Dinámicos de Respuesta Binaria con Efectos no Observables

Modelo Logit de Efectos no Observables

- Reemplazando la función de distribución acumulada de la normal estándar por la de la logística en (2) tenemos:

$$P(y_{jt} = 1 \mid X_j, c_j) = P(y_{jt} = 1 \mid x_{jt}, c_j) = \Lambda(x_{jt}\beta + c_j), t = 1, 2, \dots, T \quad (6)$$

- Manteniendo los supuestos (3) y (4) llegamos a lo que se conoce como el modelo Logit de efectos aleatorios.
- Este modelo no es tan atractivo como el modelo Probit de efectos aleatorios porque no existen estimadores fáciles de obtener.
- El modelo Probit de efectos aleatorios utiliza el hecho de que la combinación de dos funciones normales es otra normal. En el caso del Logit, $P(y_{jt} = 1 \mid X_j)$ no tiene una forma simple porque hay que integrar la función $\Lambda(X_t\beta + c)$ con respecto a la densidad de la normal $(1/\sigma_c)\phi(c/\sigma_c)$ lo que no da una forma funcional simple.

Modelo Logit de Efectos no Observables

- La mayor ventaja del modelo Logit de efectos no observables por sobre el Probit es que bajo los supuestos (6) y (3) es posible obtener estimaciones consistentes de β sin hacer supuestos acerca de la relación de c_j con x_{jt} .
- En el caso lineal, utilizamos la transformación de FE ó FD para eliminar c_j de la ecuación a estimar.
- En el caso del Logit, se puede emplear una estrategia similar. Lo que necesitamos hacer es encontrar la distribución conjunta de $y_j \equiv (y_{j1}, y_{j2}, \dots, y_{jT})'$ condicional a X_j , c_j y $n_j \equiv \sum_{t=1}^T y_{jt}$.
- Se puede verificar que esta distribución no depende de c_j tal que queda la distribución de y_j condicionada a X_j y n_j .

Modelo Logit de Efectos no Observables

- Ilustremos el caso con $T = 2$, donde n_j puede adoptar los valores 0, 1, 2. Intuitivamente la distribución condicional de $(y_{j1}, y_{j2})'$ dado n_j no puede ser informativa acerca de β cuando $n_j = 0$ ó $n_j = 2$ porque esos valores determinan completamente el resultado de y_j .
- Sin embargo para $n_j = 1$ tenemos:

$$\begin{aligned} P(y_{j2} = 1 \mid X_j, c_j, n_j = 1) &= \frac{P(y_{j2} = 1, n_j = 1 \mid X_j, c_j)}{P(n_j = 1 \mid X_j, c_j)} = \\ &= \frac{P(y_{j2} = 1 \mid X_j, c_j) P(y_{j1} = 0 \mid X_j, c_j)}{\{P(y_{j1} = 0, y_{j2} = 1 \mid X_j, c_j) + P(y_{j1} = 1, y_{j2} = 0 \mid X_j, c_j)\}} \\ &= \frac{\Lambda(x_{j2}\beta + c_j) [1 - \Lambda(x_{j1}\beta + c_j)]}{\{[1 - \Lambda(x_{j1}\beta + c_j)] \Lambda(x_{j2}\beta + c_j) + \Lambda(x_{j1}\beta + c_j) [1 - \Lambda(x_{j2}\beta + c_j)]\}} \\ &= \Lambda[(x_{j2} - x_{j1}) \beta] \end{aligned}$$

Modelo Logit de Efectos no Observables

- Similarmente, $P(y_{j1} = 1 \mid X_j, c_j, n_j = 1) = \Lambda[-(x_{j2} - x_{j1})\beta] = 1 - \Lambda[(x_{j2} - x_{j1})\beta]$
- El logaritmo de la función de verosimilitud condicional para la observación j se puede escribir como:

$$\ell_j(\beta) = 1[n_j = 1] (w_j \log \Lambda[(x_{j2} - x_{j1})\beta]) + (1 - w_j) \log \{1 - \Lambda[(x_{j2} - x_{j1})\beta]\}$$

- Donde $w_j = 1$ si $(y_{j1} = 0, y_{j2} = 1)$ y $w_j = 0$ si $(y_{j1} = 1, y_{j2} = 0)$
- El estimador de máxima verosimilitud condicional se obtiene maximizando la suma de $\ell_j(\beta)$ sobre j .
- La función indicador $1[n_j = 1]$ selecciona las observaciones para las que $n_j = 1$.
- Note que la función de verosimilitud anterior es la función estándar del Logit de corte transversal para una regresión de w_j sobre $x_{j2} - x_{j1}$ utilizando las observaciones para las que $n_j = 1$.

Modelo Logit de Efectos no Observables

- El estimador de máxima verosimilitud condicional que se obtiene en este caso se denomina **estimador Logit de efectos fijos**.
- Lo que hace este tipo de estimación es simplemente encontrar la distribución condicional, que describe la subpoblación con $n_j = 1$, que depende solo de los datos observados y de β .
- Para el caso de T general, la función de verosimilitud es un poco más complicada pero tiene un tratamiento posible.

- Primero, se tiene que:

$$P(y_{j1} = y_1, \dots, y_{jT} = y_T \mid x_j, c_j, n_j = n) = \frac{P(y_{j1}=y_1, \dots, y_{jT}=y_T \mid X_j, c_j)}{P(n_j=n \mid X_j, c_j)}$$

- Y el numerador se puede expresar como:

$$P(y_{j1} = y_1 \mid X_j, c_j) P(y_{j2} = y_2 \mid X_j, c_j) \dots P(y_{jT} = y_T \mid X_j, c_j)$$

- Usando el supuesto (3). El denominador es la parte complicada pero es fácil de describir.

Modelo Logit de Efectos no Observables

- $P(n_j = n \mid X_j, c_j)$ es la suma de las probabilidades de todos los posibles resultados de y_j tal que $n_j = n$.
- Usando la forma específica de la función logística podemos escribir:

$$\ell_j(\beta) = \log \left\{ \exp \left(\sum_{t=1}^T y_{it} x_{jt} \beta \right) \left[\sum_{a \in R_j} \exp(a_t x_{jt} \beta) \right]^{-1} \right\}$$

- Donde R_j es un subconjunto de R^T definido como $\left\{ a \in R^T : a_t \in \{0, 1\} \text{ y } \sum_{t=1}^T a_t = n_j \right\}$.
- El logaritmo de la función de verosimilitud anterior debe sumarse a través de j para luego maximizarse y obtener estimadores de β consistentes y asintóticamente normales.
- El estimador Logit de efectos fijos de β nos da el efecto de cada elemento de X_t en el logaritmo de la tasa de probabilidad.

Modelo Logit de Efectos no Observables

- Esto es,

$$\log \left\{ \frac{\Lambda(X_t\beta + c)}{1 - \Lambda(X_t\beta + c)} \right\} = X_t\beta + c$$

- Desafortunadamente, no se puede estimar el efecto parcial sobre la probabilidad de respuesta a menos que insertemos el valor de c .
- Como la distribución de c_j no está especificada es difícil saber que valor insertar.
- Además, tampoco se pueden calcular los efectos parciales promedio porque requerirían encontrar $E[\Lambda(X_t\beta + c)]$ que necesita una especificación de la distribución de c_j .

Modelos Dinámicos con Efectos no Observables

- Supongamos que la primera observación es con $t = 0$, de forma tal que y_{j0} es la primera observación de la variable y . Para $t = 1, 2, \dots, T$ estamos interesados en el modelo dinámico:

$$P(y_{jt} = 1 \mid y_{j,t-1}, \dots, y_{j0}, Z_j, c_j) = G(z_{jt}\delta + \rho y_{j,t-1} + c_j)$$

- Donde z_{jt} es un vector de variables explicativas contemporáneas, $Z_j = (z_{j1}, \dots, z_{jT})$ y G puede ser la función Probit o Logit.
- Los puntos importantes del modelo anterior son:
- Primero, los z_{jt} se asume que satisfacen un supuesto de exogeneidad estricta (condicional en c_j) dado que Z_j aparece en el conjunto de condiciones del lado izquierdo de la ecuación pero solo aparece z_{jt} en el lado derecho.
- Segundo, la probabilidad de éxito en t depende del resultado en $t - 1$ y de los efectos no observables, c_j .

Modelos Dinámicos con Efectos no Observables

- Como podríamos estimar δ y ρ en la ecuación anterior?
- Primero podemos construir la función de probabilidad conjunta de las observaciones de la muestra: $f(y_1, y_2, \dots, y_T \mid y_0, z, c; \beta)$

$$\begin{aligned} &= \prod_{t=1}^T f(y_t \mid y_{t-1}, \dots, y_1, y_0, z_t, c; \beta) \\ &= \prod_{t=1}^T G(z_t \delta + \rho y_{t-1} + c)^{y_t} [1 - G(z_t \delta + \rho y_{t-1} + c)]^{1-y_t} \quad (D.1) \end{aligned}$$

- El problema es que con T fijo, debido a la presencia del efecto no observable c , no se puede construir la función de verosimilitud que nos permita estimar β consistentemente.
- Tal como pasaba en el caso con variables estrictamente exógenas, tratar a los c_j como parámetros a ser estimados no permite obtener estimadores consistentes de δ y ρ cuando $N \rightarrow \infty$

Modelos Dinámicos con Efectos no Observables

- Lo que uno debe hacer en este caso es eliminar c vía integración como hicimos antes. El problema ahora es que para hacer esto hay que realizar algún supuesto acerca del comportamiento de la primera observación de y , y_{j0} .
- Una forma de proceder es obtener la distribución conjunta de $(y_{j1}, y_{j2}, \dots, y_{jT})$ condicional en (y_{j0}, Z_j) .
- Para obtener $f(y_1, y_2, \dots, y_T \mid y_0, z, \theta)$ necesitamos la densidad de c_j dado (y_{j0}, Z_j)
- Dada una densidad $h(c \mid y_0, z; \gamma)$ tenemos $f(y_1, y_2, \dots, y_T \mid y_0, z, \theta)$

$$= \int_{-\infty}^{\infty} f(y_1, y_2, \dots, y_T \mid y_0, z, c; \beta) h(c \mid y_0, z; \gamma) dc$$

- La integral puede ser reemplazada por una suma ponderada si la distribución de c es discreta.

Modelos Dinámicos con Efectos no Observables

- Cuando $G = \Phi$ en el modelo (D.1) una elección conveniente para $h(c \mid y_0, z; \gamma)$ es: $Normal(\psi + \xi_0 y_{j0} + Z_j \xi, \sigma_a^2)$, que sigue de escribir $c_j = \psi + \xi_0 y_{j0} + Z_j \xi + a_j$, donde $a_j \sim Normal(0, \sigma_a^2)$ e independiente de (y_{j0}, Z_j) .
- Por lo tanto se puede escribir,

$$y_{jt} = 1 [\psi + z_{jt} \delta + \rho y_{j,t-1} + \xi_0 y_{j0} + Z_j \xi + a_j + e_{jt} > 0]$$

- Tal que y_{jt} dado $(y_{j,t-1}, \dots, y_{j0}, Z_j, a_j)$ sigue un modelo Probit.

Modelos Dinámicos con Efectos no Observables

- Por lo tanto, la densidad de (y_{j1}, \dots, y_{jT}) dado (y_{j0}, Z_j) tiene la siguiente forma:
 $L(c_j, \beta) =$

$$\int_{-\infty}^{\infty} \left[\prod_{j=1}^N \prod_{t=1}^T \Phi(x_{jt}\beta + c_j)^{y_{jt}} [1 - \Phi(x_{jt}\beta + c_j)]^{1-y_{jt}} \right] \\ * (1/\sigma_c) \phi(c/\sigma_c) dc$$

- Donde $x_{jt} = (1, z_{jt}, y_{j,t-1}, y_{j0}, Z_j)$ y con a y σ_a reemplazando a c y σ_c .
- Este resultado implica que podemos utilizar el software del Probit de efectos aleatorios para estimar ψ , δ , ρ , ξ_0 , ξ , y σ_c^2 .

Supuestos

- Se tienen los siguientes supuestos:

$$P(y_{jt} = 1 \mid X_j, c_j) = P(y_{jt} = 1 \mid x_{jt}, c_j)$$

$$= \Phi(x_{jt}\beta + c_j), t = 1, 2, \dots, T. \quad (2)$$

$$y_{j1}, y_{j2}, \dots, y_{jT} \text{ son independientes condicionando en } (X_j, c_j). \quad (3)$$

$$c_j \mid X_j \sim \text{Normal}(0, \sigma_c^2) \quad (4)$$

Supuestos

- Si escribimos el supuesto (2) en forma de variable latente tenemos:
- $y_{jt}^* = x_{jt}\beta + c_j + u_{jt}$, con $y_{jt} = 1 [y_{jt}^* > 0]$ y $u_{jt} \mid X_j, c_j \sim N(0, 1)$. Además se cumple (4).
- Con estos supuestos $c_j + u_{jt} \sim N(0, 1 + \sigma_c^2)$.
- Por lo tanto: $P(y_{jt} = 1 \mid X_j) = P(c_j + u_{jt} > -x_{jt}\beta \mid X_j)$
 $= \Phi \left[x_{jt}\beta / (1 + \sigma_c^2)^{1/2} \right]$, $t = 1, 2, \dots, T$.
- Se sigue inmediatamente de la ecuación anterior que pooled Probit de y sobre X estima consistentemente:

$$\beta_c = \beta / (1 + \sigma_c^2)^{1/2}$$

References

- Chamberlain, G (1980) "Analysis of covariance with qualitative data," Review of Economic Studies, 47, pp. 225-238.
- Mundlak, Y. (1978) "On the pooling of time series and cross section data," Econometrica, 46, 69-85.

Corte Transversal

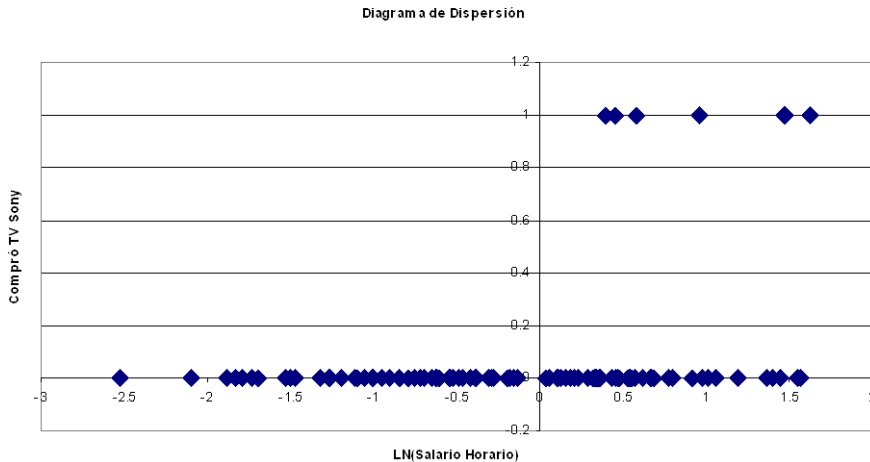
- Una variable binaria, es una variable que solo adopta dos valores (que por convención se denotan 0 y 1).
- Por ejemplo, la empresa Sony quiere saber si la venta de sus televisiones esta relacionada con el ingreso de las personas. Para esto, toma una muestra de 100 personas que en el último semestre compró un televisor y registra los valores de las siguientes variables:
- $Y = 1$ si la persona compró un televisor Sony.
- $Y = 0$ si la persona compró otra marca de TV.
- Además, registra para cada persona en la muestra el valor del salario horario.
- La empresa especifica la siguiente relación funcional:

$$Y_i = \alpha + \beta X_i + \varepsilon_i; i = 1, 2, \dots, 100.$$

- Donde X_i representa el logaritmo natural del salario horario.
- La idea de la firma es que si pudiera saber los valores de los parámetros del modelo entonces podría conocer si existe una asociación directa o inversa entre lo que gana la gente y la compra de sus televisores.
- Como primera aproximación al tema veamos un diagrama de dispersión de las variables.

Corte Transversal

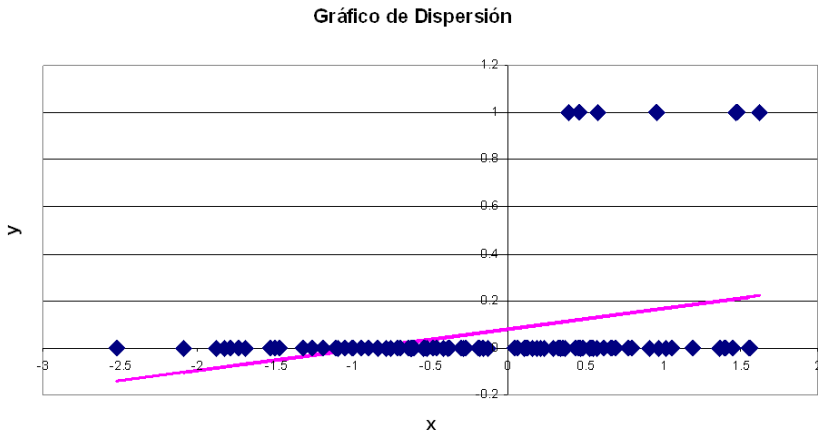
- Se tiene el siguiente diagrama de dispersión:



- Observando el gráfico de dispersión la empresa piensa que si bien hay evidencia de que la gente que compra sus televisores en general está ubicada en el sector de mayores ingresos necesita mayor información para poder segmentar el mercado.
- Esto es, si pudiera estimar los parámetros del modelo quizás pudiera distinguir a un nuevo cliente potencial.
- Utilizando el método de mínimos cuadrados clásicos, la empresa estima el modelo y obtiene los siguientes resultados: $\hat{Y}_i = 0.08 + 0.087X_i$, $i = 1, 2, \dots, 100$, con valores t 3.2 y 3.4 respectivamente.

Modelos de Respuesta Binaria

- Esto es, existe una relación directa (positiva) entre el salario y la compra de televisores Sony como se apreciaba en el gráfico de dispersión.



Corte Transversal

- Como puede observarse claramente en el gráfico anterior, el problema con esta estimación es que hay muchos valores ajustados para la variable dependiente que son NEGATIVOS!
- Esta es una consecuencia del método de estimación elegido, ya que MCC no contiene ninguna restricción que nos diga que la estimación de la variable dependiente siempre tiene que ser cero o uno.
- El problema anterior puede re-plantearse de la siguiente manera. Cuál es la probabilidad de que una persona compre un televisor Sony, dado el valor de su salario horario?
- Evidentemente, en la muestra de 100 personas es poco probable que alguien pueda responder a esa pregunta. Sin embargo, podemos observar el valor de esta probabilidad ex—post.

Corte Transversal

- Es decir, una persona ex - post o compró el televisor Sony (probabilidad de realización igual a uno) o no lo compró (probabilidad de realización igual a cero).
- Esto significa que lo que observamos es la realización de una variable no observable (la probabilidad de compra del televisor Sony).
- Observe que si la empresa pudiera conocer la probabilidad ex – ante, entonces podría segmentar el mercado con esta probabilidad.
- Es decir que a nosotros nos interesa el valor de la probabilidad ex – ante. Esto es: $Pr(y_i = 1 \mid X_i)$
- Notemos que de acuerdo al modelo especificado anteriormente tenemos:
$$Y_i = \alpha + \beta X_i + \varepsilon_i$$

- Por lo tanto tenemos que: (1) $E(Y_i | X_i) = \alpha + \beta X_i + E(\varepsilon_i | X_i) = \alpha + \beta X_i$
- Además, sabemos que por definición de esperanza matemática, la esperanza matemática condicional de una variable es la suma de cada uno de los valores que adopta la variable multiplicados por su probabilidad de ocurrencia.
- En este caso se tiene que: (2)
$$E(Y_i | X_i) = 1 * P(Y_i = 1 | X_i) + 0 * P(Y_i = 0 | X_i)$$
- $= P(Y_i = 1 | X_i)$
- Igualando las ecuaciones (1) y (2) tenemos: (3) $P(Y_i = 1 | X_i) = \alpha + \beta X_i$

Corte Transversal

- Como puede observarse en la ecuación (3) la probabilidad condicional de que el evento Y_i ocurra (en este caso la compra de la televisión Sony) dado que conocemos el valor de X_i esta expresada como una relación lineal.
- Debido a este hecho, los modelos de variable dependiente binaria reciben el nombre de Modelos de Probabilidad Lineal.
- Dado que Y_i solo puede adoptar dos valores (cero ó uno) podemos obtener la distribución de probabilidad de la variable Y_i ilustrada con la siguiente tabla a continuación:

Y_i	$Pr(Y_i X_i)$
1	$\alpha + \beta X_i$
0	$1 - (\alpha + \beta X_i)$

Corte Transversal

- Dada la distribución de la variable dependiente podemos utilizar el modelo para obtener la distribución de los errores.
- Esto es, como $Y_i = \alpha + \beta X_i + \varepsilon_i$, cuando:
- $y_i = 1 \rightarrow \varepsilon_i = 1 - \alpha - \beta X_i$ y cuando
- $y_i = 0 \rightarrow \varepsilon_i = -\alpha - \beta X_i$.
- Por lo tanto se tiene la siguiente tabla:

ε_i	$\Pr(Y_i X_i)$
$1 - \alpha - \beta X_i$	$\alpha + \beta X_i$
$-\alpha - \beta X_i$	$1 - (\alpha + \beta X_i)$

- Cuáles son los momentos de esta distribución?

$$E(\varepsilon_i | X_i) = [1 - (\alpha + \beta X_i)] (\alpha + \beta X_i) - (\alpha + \beta X_i) [1 - (\alpha + \beta X_i)] = 0$$

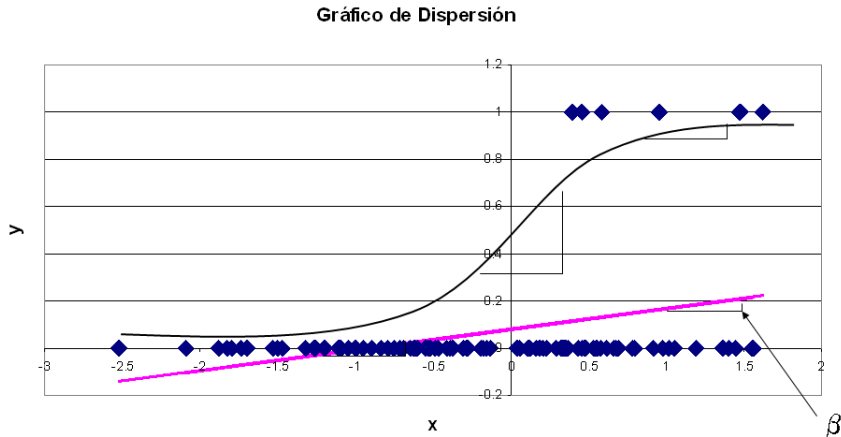
Corte Transversal

- Y por otro lado, la varianza: $Var(\varepsilon_i | X_i) = E(\varepsilon_i^2 | X_i) - E(\varepsilon_i | X_i)^2 = E(\varepsilon_i^2 | X_i) = [1 - (\alpha + \beta X_i)] * (\alpha + \beta X_i)$
- Entonces, cuando la variable dependiente es binaria, el modelo tiene heterocedasticidad.
- Esto puede generalizarse a cualquier modelo que tenga por variable dependiente una variable categórica.
- Ahora podemos resumir las características de los modelos de probabilidad lineal (MPL):
 1. Estos modelos reciben este nombre porque la variable dependiente puede interpretarse como una probabilidad.
 2. Los valores estimados, por el método de MCC, de la variable dependiente pueden caer fuera del rango $[0, 1]$ lo cual hace que la estimación sea pobre.
 3. Los errores del modelo son heterocedásticos por lo tanto MCC nos dará estimadores ineficientes

- Otro de los problemas que sufren los modelos de probabilidad lineal es el de la interpretación.
- En nuestro caso particular el coeficiente β mide cuanto afecta a la probabilidad de comprar un televisor Sony un cambio en el salario de las personas.
- Económicamente uno pensaría que este efecto debiera ser pequeño para aquellos que ganan muy poco o ganan mucho y debiera ser más grande para el resto.
- Sin embargo, en el MPL el efecto es constante e igual a β lo que no coincide con lo que se desprende de la teoría económica.
- Cómo podríamos resolver este problema de interpretación y los problemas estadísticos?

Corte Transversal

- Lo que necesitamos es una función de probabilidad que no sea lineal. Esto es:



Corte Transversal

- Las características de la curva de la figura anterior resuelve nuestros problemas ya que:
 1. Empieza en cero y termina en uno. Esto es, solo adopta valores en el intervalo $[0, 1]$.
 2. Tiene diferentes pendientes en distintos puntos. Para valores muy pequeños y muy grandes de X_i la pendiente es chica y para valores intermedios de X_i la pendiente es más grande.
- Cualquier curva de probabilidad acumulada cumple con las características antes mencionadas.
- Por lo tanto, uno puede especificar el modelo utilizando la función de probabilidad acumulada de cualquier distribución.
- Las funciones más utilizadas son la distribución Normal y la distribución Logística. [▶ Back](#)