Econometría de Datos de Panel

Maestrías en Economía y Econometría

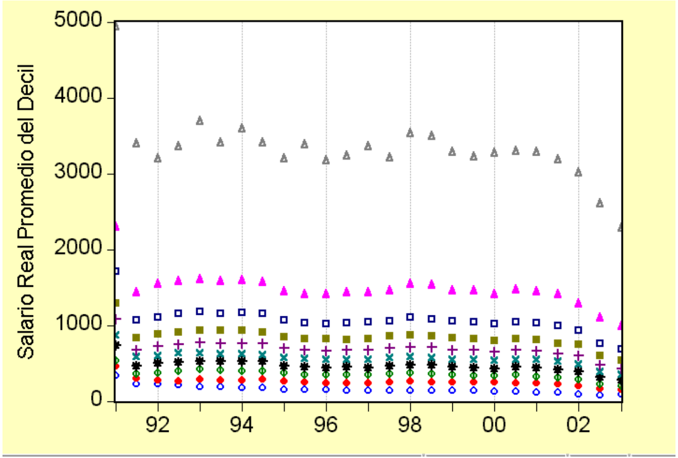
Lecture 1

Agenda

Introducción

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

Introducción



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

Convergencia en Probabilidad

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

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

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 $\{x_N : N = 1, 2, \ldots\}$ 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 $\{Z_N : N = 1, 2, \ldots\}$ 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 $g : \Re^K \to \Re^J$ una función continua en algun punto $c \in \Re^K$. Sea $\{x_N : N = 1, 2, \ldots\}$ una secuencia de vectores aleatorios $K \times 1$ tal que $x_N \xrightarrow{p} c$. Entonces $g(x_N) \xrightarrow{p} g(c)$ cuando $N \to \infty$. 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 $\{\Omega_N : N = 1, 2, \ldots\} \subset \mathfrak{F}$ se dice que ocurre con probabilidad aproximándose a uno (w.p.a 1) sí y solo sí $P(\Omega_N) \to 1$ cuando $N \to \infty$.
- Corolario 1: Sea $\{Z_N : N = 1, 2, \ldots\}$ 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 $F_N(\xi) \to F(\xi)$ cuando $N \to \infty$ para todo $\xi \in \Re$.
- 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 $g : \Re^K \rightarrow \Re^J$ 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) $z_N' V^{-1} z_N \xrightarrow{d} \chi^2_K$.
- 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 $\{w_j : j = 1, 2, \ldots\}$ una secuencia de vectores aleatorios $G \times 1$, independientes, idénticamente distribuidos tal que $E(|w_{jg}|) < \infty$, $g = 1, 2, \ldots, G$. 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 $\mu_w = E(w_j)$.

Convergencia en Distribución

- Teorema 2 (Lindeberg-Levy): Sea $\{w_j : j = 1, 2, \ldots\}$ una secuencia de vectores aleatorios $G \times 1$, independientes, idénticamente distribuidos tal que $E(|w_{jg}^2|) < \infty$, $g = 1, 2, \ldots, G$ 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 $B = Var(w_j) = E(w_j w_j')$.

Modelos de Datos de Panel Lineales

- Supuestos: se dispone de una muestra aleatoria de la población. Esto es, tenemos observaciones de CORTE TRANSVERSAL $\{(X_j, y_j) : j = 1, 2, \ldots, N\}$ 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:

$$
\begin{aligned}
y_{jt} &= x_{jt}\beta + u_{jt}, \quad j = 1, 2, \ldots, N; t = 1, 2, \ldots, T && (1) \\
y_j &= X_j\beta + u_j, \quad j = 1, 2, \ldots, N && (2)
\end{aligned}
$$

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:

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

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

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

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

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 $E(x_{jt}'u_{jt}) = 0, \quad t = 1, 2, \ldots, T$.
- Note que la ecuación anterior NO impone exogeneidad estricta. Bajo el supuesto 1, el vector β satisface:

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

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

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

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

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

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

- 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

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

- 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:

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

- 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:
$\frac{1}{\sqrt{N}} \sum_{j=1}^{N} \sum_{t=1}^{T} x'_{jt} u_{jt} \xrightarrow{d}$ Normal$(0, B)$, donde:
$B \equiv E\left(\sum_{t=1}^{T} x'_{jt} u_{jt} u'_{jt} x_{jt}\right) = Var(\sum_{t=1}^{T} x'_{jt} u_{jt})$

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

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

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

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

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

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

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

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

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

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

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

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

- Entonces,

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

Modelos de Datos de Panel Lineales

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

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

- 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 esta 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.

- 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:

$$W = (R\hat{\beta} - r)'[R\widehat{V}R']^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi^2_Q \tag{14}$$

- 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) $E(u^2_{jt}x'_{jt}x_{jt}) = \sigma^2 E(x'_{jt}x_{jt})$, $t = 1, 2, \ldots, T$ y $\sigma^2 = E(u^2_{jt})$ para todo t. (b) $E(u_{jt}u_{js}x'_{jt}x_{js}) = 0$, $t \neq s$, $t, s = 1, 2, \ldots, T$.

Modelos de Datos de Panel Lineales

- El supuesto 3 implica que,

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

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

- El estimador apropiado es entonces,

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

- 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′. $E(X_j \bigotimes u_j) = 0$. 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 $E(u_j^* u_j^{*\prime}) = I_T$.

Modelos de Datos de Panel Lineales

- Aplicando OLS a (16) obtenemos:

$$
\begin{aligned}
\tilde{\beta} &= \left(\frac{1}{N}\sum_{j=1}^{N} X_j^{*\prime} X_j^{*}\right)^{-1}\left(\frac{1}{N}\sum_{j=1}^{N} X_j^{*\prime} y_j^{*}\right) \\
&= \left(\frac{1}{N}\sum_{j=1}^{N} X_j'\Omega^{-1}X_j\right)^{-1}\left(\frac{1}{N}\sum_{j=1}^{N} X_j'\Omega^{-1}y_j\right) \qquad (17) \\
&= \beta + \left(\frac{1}{N}\sum_{j=1}^{N} X_j'\Omega^{-1}X_j\right)^{-1}\left(\frac{1}{N}\sum_{j=1}^{N} X_j'\Omega^{-1}u_j\right) \qquad (18)
\end{aligned}
$$

- Por la ley de los grandes números (WLLN) $\left(\frac{1}{N}\sum_{j=1}^{N} X_j'\Omega^{-1}X_j\right) \xrightarrow{p} E\left(X_j'\Omega^{-1}X_j\right) \equiv A$

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)

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

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

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

Modelos de Datos de Panel Lineales

- Usando la ecuación (18) tenemos,

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

- Como vimos, por la WLLN y el Supuesto 1', $E\left(X_j'\Omega^{-1}u_j\right) = 0$ 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 $B = E(X_j'\Omega^{-1}u_ju_j'\Omega^{-1}X_j)$.

Modelos de Datos de Panel Lineales

- Usando los resultados anteriores, tenemos

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

- Por lo tanto $\mathrm{AVar}(\tilde{\beta}) = A^{-1} B A^{-1} / N$.
- 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.

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

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

- El segundo término del lado derecho puede escribirse como

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

- El último término del lado derecho puede escribirse como:

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

Modelos de Datos de Panel Lineales

- Por lo tanto

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

- El estimador FGLS de β es entonces:

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

Modelos de Datos de Panel Lineales

- De la última expresión tenemos

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

- Ahora,

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

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

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

- 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.
- Resúmen: Bajo los supuestos 1' y 2',

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

- 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,

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

Modelos de Datos de Panel Lineales

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

$$\widehat{V} = \widehat{\text{AVar}}(\hat{\hat{\beta}}) = \tilde{A}^{-1}\tilde{B}\tilde{A}^{-1}/N =$$

$$\left(\sum_{j=1}^{N} X_j'\tilde{\Omega}^{-1}X_j\right)^{-1}\left(\sum_{j=1}^{N} X_j'\tilde{\Omega}^{-1}\tilde{u}_j\tilde{u}_j'\tilde{\Omega}^{-1}X_j\right)\left(\sum_{j=1}^{N} X_j'\tilde{\Omega}^{-1}X_j\right)^{-1} \tag{22}$$

- 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'. $E(X_j'\Omega^{-1}u_ju_j'\Omega^{-1}X_j) = E(X_j'\Omega^{-1}X_j)$, con $\Omega = E(u_ju_j')$.
- Bajo los supuestos 1', 2' y 3', la varianza asintótica del estimador FGLS es:

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

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

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

- Los errores estándar asintóticos de los coeficientes estimados se obtienen en forma usual utilizando la raiz 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 $E(e_{jt}|X_{jt}, u_{jt-1}, \ldots) = 0$.
- 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) $\text{LM} = \text{Obs} \times R^2_{ax} \sim \chi^2_1$.
  (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 $E(u_{jt}|x_{jt}) = 0, \quad t = 1, 2, \ldots, T.$
- La hipótesis nula es entonces $E(u_{jt}^2|x_{jt}) = \sigma^2, \quad t = 1, 2, \ldots, T.$
- 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) $\text{LM} = N \times T \times R_{ax}^2 \sim \chi_Q^2$.
  (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 $x = (x_1, x_2, \ldots, x_k)$ 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 $\mathrm{Cov}(x_j, c) \neq 0$ 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:

$$y_{it} = c_i + x_{it}\beta + u_{it} \tag{24}$$

  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: $\text{Cov}(x_{it}, c_i) = 0, t = 1, 2, \ldots, T$.
- 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

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

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

- 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:

$$E(x_{is}' u_{it}) = 0, \quad s, t = 1, 2, \ldots, T \tag{26}$$

Exogeneidad Estricta

- Este supuesto de exogeneidad es mucho más fuerte que el de ausencia de correlación contemporánea
- $E(x_{it}' u_{it}) = 0, \quad t = 1, 2, \ldots, T.$
- 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:

$$y_{it} = x_{it}\beta + v_{it} \quad t = 1, 2, \ldots, T \tag{27}$$

  donde $v_{it} = c_i + u_{it}$, $t = 1, 2, \ldots, T$ 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 $E(x_{it}' v_{it}) = 0, \quad t = 1, 2, \ldots, T.$

Modelos de Panel Lineales

- La última condición implica que estamos asumiendo que:
- $E(x_{it}'u_{it}) = 0$ y $E(x_{it}'c_i) = 0$, $t = 1, 2, \ldots, T$.
- Note que el supuesto restrictivo aqui 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

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 $X_i = (x_{i1}, x_{i2}, \cdots, x_{iT})$
- 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:

$$y_{it} = x_{it}\beta + v_{it} \tag{28}$$

Modelo de Efectos Aleatorios

- En (28) $E(v_{it}|X_i) = 0, \quad t = 1, 2, \ldots, T$.
- Note que esta última ecuación implica que $\{x_{it}, \ t = 1, 2, \ldots, T\}$ 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.

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

  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) $E(u_{it}^2) = \sigma_u^2, \quad t = 1, 2, \ldots, T.$
  (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:

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

- 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:

$$\widehat{\Omega} = \hat{\sigma}_c^2 J_T J_T' + \hat{\sigma}_u^2 I_T \tag{30}$$

- El estimador FGLS que usa (30) se conoce como estimador de efectos aleatorios.
- $$\hat{\beta}^{RE} = \left(\sum_{i=1}^{N} X_i' \widehat{\Omega}^{-1} X_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' \widehat{\Omega}^{-1} y_i \right)$$
- El estimador anterior es consistente bajo RE.1 y RE.2.

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

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 $\widehat{\Omega}$ dada por (30) en lugar de $\tilde{\Omega}$.

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

- 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:

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

Donde $\hat{\hat{v}}_{it}$ 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:

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

- 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 $\hat{v}_i = y_i - X_i \hat{\beta}^{RE}, \quad i = 1, 2, \ldots, N$

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

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

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

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

Modelo de Efectos Aleatorios

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

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

  donde $\hat{\hat{v}}_i$ 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 $E(v_i v_i' | X_i) = \Omega$, 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

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

  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.

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

- 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:

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

- 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

Modelo de Efectos Fijos

- Consideremos nuevamente el modelo de componentes no observados

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

donde:

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

- 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 $cov(c_i, x_{it}) \neq 0$ 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: $E(u_{it}|X_i, c_i) = 0, \quad t = 1, 2, \ldots, T$
- 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 $t = 1, 2, \ldots, T$ para obtener la ecuación de corte transversal:

$$\bar{y}_i = \bar{x}_i \beta + c_i + \bar{u}_i \tag{39}$$

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

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

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:

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

- 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 $\widehat{\beta}^{FE}$ 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, $\widehat{\beta}^{FE}$ es:

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

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: 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 $\mathrm{Var}(u_i | X_i, c_i) = \sigma_u^2 I_T$.
- 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

$$\ddot{y}_{it} = \ddot{x}_{it}\beta + \ddot{u}_{it} \tag{43}$$

- 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:

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

$$\sqrt{N}(\widehat{\beta}^{FE} - \beta) = \left(N^{-1} \sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i\right)^{-1} \left(N^{-1/2} \sum_{i=1}^{N} \ddot{X}_i' u_i\right) \tag{44}$$

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:

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

- 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.

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

Modelo de Efectos Fijos

- Los errores estándar asintóticos se obtienen con la raiz 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:

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

- 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:

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

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 útima 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:

$$F = \frac{(\mathrm{RSS}_r - \mathrm{RSS}_u)/Q}{\mathrm{RSS}_u/[N(T-1)-K]} \xrightarrow{d} F_{Q,N(T-1)-K}$$

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: 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, y x_{it} con $t = 1, 2, \ldots, T$; $j = 1, 2, \ldots, N$.
- 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,

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

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

- Re-escribiendo (48) como

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

- Obtenemos,

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

- 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.

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

- Los errores estándar de los estimadores de FE se obtienen de la raiz 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:

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

- 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

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

Modelo de Efectos Fijos

- Un estimador consistente de Ω es:

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

- El estimador de FE por GLS se define como:

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

donde las variables están definidas sin el último período temporal.

- Para obtener consistencia necesitamos una nueva condición de rango:
- Supuesto FEGLS.2: rango$E(\ddot{X}_i'\Omega^{-1}\ddot{X}_i) = K$.
- 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 $t-1$,

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it}\beta + u_{it} &(53)\\
y_{it-1} &= \alpha_i + x_{it-1}\beta + u_{it-1} &(54)
\end{aligned}
$$

- Restando miembro a miembro,

$$
\begin{aligned}
y_{it} - y_{it-1} &= (x_{it} - x_{it-1})\beta + (u_{it} - u_{it-1}) &(55)\\
\Delta y_{it} &= \Delta x_{it}\beta + \Delta u_{it} &(56)
\end{aligned}
$$

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

Modelo de Efectos Fijos

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

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

donde

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

Modelo de Efectos Fijos

- Note que para que este estimador sea consistente necesitamos que se cumpla que:
  Supuesto $FD.1: \ E(u_{it}|X_i, c_i) = 0, \ \ t = 1, 2, \ldots, T$
- El supuesto $FD.1$, es igual a FE.1 y a RE.1(a).
- Bajo supuesto $FD.1$, POLS en (56) será consistente porque $E(\Delta x'_{it}\Delta u_{it}) = 0, t = 2, \ldots, T$, 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 $1, \ T+1, 2T+1, \ \ldots, \ (N-1)T+1$ 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: $E(Du_iu_i'D'|X_i, c_i) = DE(u_iu_i'|X_i, c_i)D' = \sigma_u^2 DD'$
- Lo que muestra que los errores estarán correlacionados para períodos adjacentes.
- En este caso, argumentos estándares de GLS nos darán el siguiente estimador óptimo:

$$\widehat{\beta}^{FDGLS} = \left(\sum_{i=1}^{N} X_i' D'(DD')^{-1} DX_i\right)^{-1} \left(\sum_{i=1}^{N} X_i' D'(DD')^{-1} Dy_i\right) \qquad (59)$$

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: $\widehat{\beta}^{FDGLS} = \widehat{\beta}^{FE}$
- 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 $e_{it} = \Delta u_{it}, \;\; t = 2, \ldots, T$ (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:

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

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

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

- 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:

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

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:

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

- El modelo transformado es entonces:

$$y_{it}^{*} = x_{it}^{*}\beta + u_{it}^{*} \tag{61}$$

- 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:

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

con

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

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:

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

- 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 $\widehat{\beta}^{FE} \equiv \widehat{\beta}^{FDGLS} \equiv \widehat{\beta}^{HT}$

Modelo de Efectos Fijos

- El último punto implica que necesitamos:
- Supuesto HT.1: $E(u_{it}|X_i, c_i) = 0, \quad t = 1, 2, \ldots, T$
- 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 $\widehat{\beta}^{HT}$ 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

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

Modelo de Efectos Fijos

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

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

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

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

- Los errores estándar asintóticos se obtienen con la raiz 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:

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

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

- 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.

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

- Los errores estándar de los estimadores de HT se obtienen de la raiz 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:

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

- 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:

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

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 $\widehat{\delta}^{RE}$ al vector de estimadores de RE sin los coeficientes que acompañan a variables constantes en el tiempo; y por $\widehat{\delta}^{FE}$ 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:

$$H = (\widehat{\delta}^{FE} - \widehat{\delta}^{RE})'[\widehat{Var(\delta)}^{FE} - \widehat{Var(\delta)}^{RE}]^{-1}(\widehat{\delta}^{FE} - \widehat{\delta}^{RE}) \Longrightarrow \chi^2_M \tag{69}$$

- 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:

$$t = \frac{(\widehat{\delta_1}^{FE} - \widehat{\delta_1}^{RE})}{\sqrt{[\widehat{Var(\delta_1)}^{FE} - \widehat{Var(\delta_1)}^{RE}]}} \Longrightarrow N(0, 1) \tag{70}$$

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:

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

- La inclusión de α_i controla por factores no obervables 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:

$$\begin{aligned}
y_{it} &= \alpha_i + \lambda_t + x_{it}\beta + u_{it} && (71)\\
\bar{y}_i &= \alpha_i + \bar{\lambda} + \bar{x}_i\beta + \bar{u}_i && (72)\\
\bar{y}_t &= \bar{\alpha} + \lambda_t + \bar{x}_t\beta + \bar{u}_t && (73)\\
\bar{y} &= \bar{\alpha} + \bar{\lambda} + \bar{x}\beta + \bar{u} && (74)
\end{aligned}$$

Extensión del Modelo de Efectos Fijos

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

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

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