Econometría de Datos de Panel

Maestrías en Economía y Econometría

Lecture 3

Agenda

Agenda

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,

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

  donde x_{it} es $1 \times K$ y β es $K \times 1$, y $Cov(x_{it}, c_i) \neq 0$. Asumimos que hay disponibles N observaciones de corte transversal y que la teoría asintótica relevante es con $N \longrightarrow \infty$.
- 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 $s_i \equiv (s_{i1}, s_{i2}, \ldots, s_{iT})'$ 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 $(x_i; y_i; s_i) : i = 1; 2; \ldots; N$ 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,

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

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) $E(u_{it}|x_i; s_i; c_i) = 0$, $t = 1; 2; \ldots; T$; (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 $E(u_{it}|x_i; c_i) = 0 \ \forall \, t$.
- 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 procedimiento de inferencia estándar de FE son válidos. En particular bajo FEUP.1 (a) y FEUP.1 (c),

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

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

- Rejajar 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.

Agenda

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. (Incidentalemente, 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

$$y_{it} = x_{it}\beta + c_i + u_{it}, \quad t = 1, 2, \ldots, T \tag{3}$$

▸ 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

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

- El mecanismo de selección descripto 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 $\ddot{\hat{\lambda}}_{it} = \hat{\lambda}_{it} - T_i^{-1} \sum_{r=1}^{T} s_{ir} \hat{\lambda}_{ir}$ en la estimación de FE usando solo aquellas observaciones para las que $s_{it} = 1$,

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

- 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 Incidental

- 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 $\overline{\mathit{lnyit}}$=mean(lnyit)
  probit sit lnyit $\overline{\mathit{lnyit}}$
  predict sitf, xb
  gen lambdait=normden(sitf)/norm(sitf)
  areg lncit lnyit lambdait, absorb(firm) robust
  outreg using itrunc, 3aster

  type itrunc.out

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

Agenda

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) $E(u_{it}|x_i; v_{it}) = E(u_{it}|v_{it}) = \rho_t v_{it}, t = 1; \ldots; T$; 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,

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

- Tomando esperanzas condicionales en (3),

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

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

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

Truncamiento Incidental: Estimación

- La última ecuación lleva al siguiente procedimiento,
  1. 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. $h_{it} \equiv (1\, \bar{x}_i\, x_{it})$; $\hat{\mu} = (\hat{\eta}_0\, \hat{\eta}'\, \hat{\delta}')'$.
  2. Para $s_{it} = 1$ defina el vector $1 \times (1 + 2K + T)$,
  $$\hat{w}_{it} \quad = \quad (1, \underbrace{\bar{x}_i}_{K}, \underbrace{x_{it}}_{K}, \underbrace{0, \ldots, 0, \hat{\lambda}_{it}, 0, \ldots, 0}_{T}),$$
  y obtenga $\hat{\theta} = (\hat{\pi}_0, \hat{\pi}', \hat{\beta}', \hat{\gamma}')'$ como el estimador pooled OLS en
  $$y_{it} = \hat{w}_{it}\theta + \text{error}_{it}, \quad s_{it} = 1.$$

Truncamiento Incidental: Inferencia

- 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 $\hat{\theta}$, 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

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

     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.

- Continuación

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

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

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

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

Truncamiento Incidental: Inferencia

. (cont.) Para cada i defina,

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

- 5. Defina,

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

Entonces,

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

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. ▸ Go
- 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:

$$y_{jt} = x^1_{jt}\beta + c_j + u_{jt}, \quad t = 1, 2, ..., T \tag{5}$$

- La variable y_{jt} se observa siempre que h_{jt} se observa. Para cada $t = 1, 2, ..., T$, definamos la siguiente variable censurada:
  $h_{jt} = max(0, h^*_{jt})$, se observa, y

$$h^*_{jt} = \delta_{t0} + x_{j1}\delta_{t1} + ... + x_{jT}\delta_{tT} + v_{jt}, \tag{6}$$

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 descripto 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) $E(u_{jt} \mid X_j, v_{jt}) = E(u_{jt} \mid v_{jt}) = \rho_t v_{jt},\ t = 1, 2, ..., T$; 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 $X_j = (1, x_{j1}, x_{j2}, ..., x_{jt})$ y $\delta_t = (\delta_{t0}, \delta_{t1}, ..., \delta_{tT})$. Para $s_{jt} = 1$, defina

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

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

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

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, $Av\hat{a}r(\hat{\theta})$, 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 $1 \times (1 + TK)$, $-Xj$, 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” $\{\hat{r}_{j1}, \hat{r}_{j2}, ..., \hat{r}_{jT}\}$.

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 $j = 1, 2, ..., N$ 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:

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

- 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 estríctamente 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 $N \to \infty$.
- 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:

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

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

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

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

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:

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

- 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:
  $\xi_j = \eta_0 + x_{j1}\eta_1 + \ldots + x_{jT}\eta_T + \alpha_j$
- (b) los errores de la ecuación de selección tienen distribución normal, $v_{jt} = a_{jt} + \xi_j \sim N\left(0, \sigma_t^2\right)$.
- (c) los errores $\left[\left(u_{jt} - u_{js}\right), v_{jt}, v_{js}\right]$ 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:
$E\left(u_{jt} - u_{js} \mid X_j, v_{jt} \geq -H_{jt}, v_{js} \geq -H_{js}\right) = \sigma_{(u_j - u_s)(v_t/\sigma_t)}\lambda_{jts} + \sigma_{(u_j - u_s)(v_s/\sigma_s)}\lambda_{jts},$
- Donde $H_{j\tau} = X_{j\tau}\delta + E\left(\xi_j \mid X_j\right)$ para $\tau = t, s$, es la forma reducida de los indicadores de selección para los períodos t y s.
- Además, $\lambda_{jts} = \phi(M_{jt})\Phi\left(M^*_{jts}\right)/\Phi_2(M_{jt}, M_{js}, \rho_{ts})$, y
$\lambda_{jst} = \phi(M_{js})\Phi\left(M^*_{jtt}\right)/\Phi_2(M_{jt}, M_{js}, \rho_{ts})$.
- Donde, $M_{jt} = \frac{H_{jt}}{\sigma_t}$, $M_{jt} = \frac{H_{js}}{\sigma_s}$, $M^*_{jts} = (M_{js} - \rho_{ts}M_{jt})/(1 - \rho^2_{ts})^{1/2}$ y
$M^*_{jst} = (M_{jt} - \rho_{ts}M_{js})/(1 - \rho^2_{ts})^{1/2}$

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 $\phi(.)$ es la función de densidad y $\Phi(.)$ y $\Phi_2(.)$ son las funciones acumuladas de la distribución normal estándar univariante y bivariante, respectivamente
- Por lo tanto, la ecuación a estimar queda:

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

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

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

- 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 $\left(\begin{array}{c} T \\ 2 \end{array}\right)$ 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 $\left(\begin{array}{c} T \\ 2 \end{array}\right)$ 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.

Agenda

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.

$$y_{jt} = x_{jt}\beta + c_j + u_{jt}, \tag{10}$$

  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.
- $$\Delta y_{jt} = \Delta x_{jt}\beta + \Delta u_{jt}, \quad t = 2, \ldots, T \tag{11}$$

Attrition: Corrección

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

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

  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 estríctamente 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

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

- Entonces,

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

- 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}, $d2_t\hat{\lambda}_{jt}$, $\ldots$, $dT_t\hat{\lambda}_{jt}$, $t = 2, \ldots, T$, 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 $H_0 : \rho_t = 0,\ t = 2, \ldots, T$ 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,

$$y_{it} = x_{it}\beta + v_{it} \tag{15}$$

  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}: $E(v_{it}|z_{it}) = 0$, $t = 1, 2, \ldots, T$.
- 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 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,

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

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

- La ecuación (16) se puede escribir como,

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

- Para T fijo y $N \longrightarrow \infty$ 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) $E(v_{it}|z_{it}, s_{it}) = 0,\ t = 1, 2, \ldots, T$; (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,

$$y_{it} = x_{it}\beta + c_i + u_{it} \tag{18}$$

  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 estríctamente 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 $\{u_{ir} : r = 1, 2, \ldots, T\}$.
- 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

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

- Usando álgebra,

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

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

- Denote por $z_i = (z_{i1}, \ldots, z_{iT})$ y $s_i = (s_{i1}, \ldots, s_{iT})$. 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) $E(u_{it}|z_i, s_i, c_i) = 0$, $t = 1, 2, \ldots, T$;
  (iii) rango $E\left(\sum_{t=1}^{T} s_{it}\ddot{z}_{it}'\ddot{x}_{it}\right) = K$;
  (iv) 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 $N \longrightarrow \infty$.
- El Supuesto FE2SLS.1 (ii) siempre se satisface si los z_{it} son estríctamente 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:
$$y_{it1} = x_{it}\beta_1 + c_{i1} + u_{it1}, \quad t = 1, \ldots, T \tag{21}$$
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}, estríctamente 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},

$$s^*_{it2} = z_{it}\delta_2 + c_{i2} + u_{it2}, \quad t = 1, \ldots, T \tag{22}$$

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

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

$$s_{it2} = 1[s^*_{it2} > 0] = 1[z_{it}\delta_2 + c_{i2} + u_{it2} > 0], \tag{23}$$

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

- Para derivar el test asumimos que

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

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),

$$\begin{aligned} c_{i2} &= \bar{z}_i \xi_2 + a_{i2}, && (25) \\ a_{i2} | z_i &\sim \text{Normal}(0, \tau_2^2), \quad t = 1, \ldots, T && (26) \end{aligned}$$

  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,

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

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,

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

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

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

donde, por ahora ρ_1 es constante en el tiempo.

- Bajo el supuesto anterior:

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

Contraste por Sesgo de Selección bajo Truncamiento Incidental

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

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

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

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

- 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 bivariante truncada obtenemos el cálculo Probit usual,

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

  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:

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

   Usando los resultados de la estimación construya la inversa de los cocientes de Mills: $\widehat{\lambda}_{it2} \equiv \lambda(z_{it}\hat{\delta}_{t2} + \bar{z}_i\hat{\xi}_{t2})$.
2. Para la muestra observada, estime (31) usando FE-2SLS, reemplazando $E(v_{it2}|z_i, s_{it2})$ por $\widehat{\lambda}_{it2}$.

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 $\widehat{\lambda}_{it2}$ 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 esta 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,

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

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:

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

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) $E(v_{it1}|z_i, v_{it2}) \equiv E(a_{i1} + u_{it1}|z_i, v_{it2}) = E(a_{i1} + u_{it1}|v_{it2}) = \gamma_{t1} v_{it2}, \;\; t = 1, \ldots, T$

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- Usando (iii) y (iv) tenemos,

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

- Escribiendo el modelo para la muestra que observamos queda,

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

- Esto significa que podemos estimar β_1, ξ_1 y $(\gamma_{11}, \gamma_{21}, \ldots, \gamma_{T1})$ usando Pooled 2SLS una vez que reemplazamos λ_{it2} con $\widehat{\lambda}_{it2}$.
- 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 $1, z_{it}, \bar{z}_i$, $i = 1, \ldots, N$, y obtenga la inversa de los cocientes de Mills, $\widehat{\lambda}_{it2}$.
2. Para la muestra de datos observados, estime la ecuación (38) con λ_{it2} reemplazada por $\widehat{\lambda}_{it2}$ por Pooled 2SLS usando $1, z_{it}, \bar{z}_i, \widehat{\lambda}_{it2}$ 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: $\hat{w}_{it} = (x_{it}, \bar{z}_i, 0, \ldots, 0, \widehat{\lambda}_{it2}, 0, \ldots, 0)$ y $\hat{h}_{it} = (z_{it}, \bar{z}_i, 0, \ldots, 0, \widehat{\lambda}_{it2}, 0, \ldots, 0)$, respectivamente. Sin el "^" estos vectores están evaluados en los parámetros poblacionales.

- El vector de parámetros de la ecuación de interés es: $\theta = (\beta_1', \xi_1', \gamma_{11}', \ldots, \gamma_{T1}')'$.

- El vector de parámetros de la ecuación de selección es: $\pi_t = (\delta_{t2}', \xi_{t2}')'$, y $\pi = (\pi_1', \pi_2', \ldots, \pi_T')'$.

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- El estimador de Pooled 2SLS es,

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

- 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

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

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 $E(e_{it1}|h_{it}, s_{it2}) = 0$,

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

donde $F = E[\sum_{t=1}^{T} s_{it2} h'_{it} (\theta' \bigtriangledown_{\pi} w'_{it})]$, $\bigtriangledown_{\pi} w'_{it}$ 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.

- Combinando (41) con (40) tenemos,

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

- 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:

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

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

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

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

Corrección del Sesgo de Selección bajo Truncamiento Incidental

- En la expresión anterior para cada $(i,\ t)$, $\hat{\theta}' \bigtriangledown_\pi \hat{w}'_{it} = (0, \ldots, 0, -\hat{\gamma}_{t1} q_{it} \hat{\lambda}_{it2}(q_{it}\hat{\pi}_t + \hat{\lambda}_{it2}), 0, \ldots, 0,)$ 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,

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

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

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

- En la expresión anterior

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

  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}, $i = 1, \ldots, N$.
- 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:

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

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

$$
\begin{aligned}
(y_i \mid y_i \text{ es observada}) &= E\left[y_i \mid z_i^* > 0\right] && (50)\\
&= E\left[y_i \mid u_i > -\gamma' w_i\right] && (51)\\
&= \beta' x_i + E\left[\varepsilon_i \mid u_i > -\gamma' w_i\right] && (52)\\
&= \beta' x_i + \rho \sigma_\varepsilon \lambda_i(\alpha_u) && (53)\\
&= \beta' x_i + \beta_\lambda \lambda_i(\alpha_u), && (54)
\end{aligned}
$$

- 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:

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

- 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 $\phi(.)$ y por $\Phi(.)$ 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:

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

- 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:

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

- Las condiciones de primer orden son:

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

Modelo Probit

- La matriz de información de Fisher es:

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

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

Go Back

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:

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

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

Modelo Tobit

- 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, $\gamma_i = \frac{\phi_i}{1-\Phi_i}$,
  $Y_1' = (y_i, y_2, ..., y_{N_1})$,
  $X_1' = (x_1, x_2, ..., x_{N_1})$,
  $X_0' = (x_{N_1+1}, x_{N_1+2}, ..., x_N)$ y
  $\gamma_0' = (\gamma_{N_1+1}, \gamma_{N_1+2}, ..., \gamma_N)$.
- 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
  $Prob(y_i > 0) f_i(y_i \mid y_i > 0) = F_i \frac{f(y_i - \beta' x_i, \sigma^2)}{F_i} = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-(1/2\sigma^2)(y_i - \beta' x_i)^2}$

Modelo Tobit

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

Modelo Tobit

- Usando las siguientes expresiones:
- $\frac{\partial F_i}{\partial \beta} = f_i x_i$
- $\frac{\partial F_i}{\partial \sigma^2} = -\frac{1}{2\sigma^2}\beta' x_i f_i$
- $\frac{\partial f_i}{\partial \beta} = -\frac{1}{\sigma^2}\beta' x_i f_i x_i$
- $\frac{\partial f_i}{\partial \sigma^2} = \frac{(\beta' x_i)^2 - \sigma^2}{2\sigma^4} f_i$

Modelo Tobit

- Las condiciones de primer orden quedan:
  $\frac{\partial logL}{\partial \beta} = -\sum_{\forall y_i=0} \frac{f_i x_i}{1-F_i} + \frac{1}{\sigma^2} \sum_{\forall y_i>0} \left(y_i - \beta' x_i\right) x_i = 0$ y
  $\frac{\partial logL}{\partial \sigma^2} = \frac{1}{2\sigma^2} \sum_{\forall y_i=0} \frac{\beta' x_i f_i}{1-F_i} - \frac{N_1}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{\forall y_i>0} (y_i - \beta' x_i)^2 = 0$
- 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:

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

Modelo Tobit

- Las condiciones de segundo orden son:
- $\frac{\partial^2 logL}{\partial\beta\partial\beta'} = -\sum_{\forall y_i=0} \frac{f_i}{(1-F_i)^2}\left[f_i - \frac{1}{\sigma^2}(1-F_i)\beta' x_i\right] x_i x_i' - \frac{1}{\sigma^2}\sum_{\forall y_i>0} x_i x_i'$
-

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

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

Modelo Tobit

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

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

- Por lo tanto

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

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 $j = 1, 2, ..., N$.
- Como $z_j = 1$ sí y solo sí $y_{ij} = 1$ e $y_{2j} = 1$, la distribución de z_j es,
- $p_j = Pr(z_j = 1) = Pr(y_{ij} = 1, y_{2j} = 1) = F(x_j\beta_1, x_j\beta_2; \rho)$,
- $1 - p_j = Pr(z_j = 0) = Pr(y_{ij} = 0$ 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 $F(.)$ denota la distribución normal estándar bivariada.

Modelo Probit Bivariante

- El logaritmo de la función de verosimilitud es:

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

- 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

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

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

- 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_1y 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

La Matriz de Varianzas

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

La Matriz de Varianzas

- Donde,

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

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

La Matriz de Varianzas

- La estimación en dos pasos es como sigue. Primero estime $\overline{\omega}_{ts} = (\delta_t', \delta_s', \rho_{ts})'$ con $\hat{\overline{\omega}}_{ts} = \left(\hat{\delta}_t', \hat{\delta}_s', \hat{\rho}_{ts}\right)'$
- 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 $\Delta x_{jts}^1 = (x_{jt}^1 - x_{js}^1)$ y $\left(\hat{\lambda}_{jts}, \hat{\lambda}_{jst}\right)$.

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 $R_{jts} \equiv \left(\Delta x^1_{jts}, \lambda_{jts}, \lambda_{jst}\right)'$ y las condiciones de primer orden en el segundo paso son:
- $\frac{1}{N}\sum_j s_{jt}s_{js}\left\{\Delta y_{jts} - \Delta x^1_{jts}\hat{\beta} - \hat{\ell}_{ts}\hat{\lambda}_{jts} - \hat{\ell}_{st}\hat{\lambda}_{jst}\right\}R_{jts} = 0$

La Matriz de Varianzas

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

$$\begin{aligned}
\sqrt{N}\left(\widehat{\varpi}_{ts} - \overline{\omega}_{ts}\right) & \xrightarrow{p} \frac{1}{\sqrt{N}}\sum_{j} I^{-1}_{\overline{\omega}_{ts}} \left(\begin{array}{c} X_j' \left(q_{jt}\phi_{jt}\Phi_{jts}/\Phi_{2,jts}\right) \\ X_j' \left(q_{js}\phi_{st}\Phi_{jts}/\Phi_{2,jts}\right) \\ q_{jt}q_{js}\phi_{2,jts}/\Phi_{2,jts} \end{array} \right) \\
& \equiv \frac{1}{\sqrt{N}}\sum_{j} \Lambda_j
\end{aligned}$$

La Matriz de Varianzas

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

La Matriz de Varianzas

- Además,

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

- donde,

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

La Matriz de Varianzas

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

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

- 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)

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

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

$$
\begin{aligned}
y_1 &= x_1\beta_1 + u_1 && (57)\\
y_2 &= \max(0; x\delta_2 + v_2) && (58)
\end{aligned}
$$

  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) $v_2 \sim Normal(0; \tau_2^2)$; 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

$$E(y_1|x; v_2; s_2 = 1) = x_1\beta_1 + \gamma_1 v_2 \tag{59}$$

- 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 $i = 1, 2, \ldots N_1$), define

$$\hat{v}_{i2} = y_{i2} - xi\hat{\delta}_2 \tag{60}$$

  (b) Using observations for which $y_{i2} > 0$, estimate $(\beta_1; \gamma_1)$ by the OLS regression: y_{i1} on x_{i1}, and $\hat{v}_{i2}$ $i = 1, 2, \ldots N_1$
- This regression produces consistent and $\sqrt{N}$ asymptotically normal estimators of $(\beta_1; \gamma_1)$.

Go Back