

Truncamiento y Selección

Microeconomía Cuantitativa

R. Mora

Departamento de Economía
Universidad Carlos III de Madrid

Esquema

- 1 Introducción
- 2 Truncamiento
- 3 MCO y el modelo de Heckman
- 4 El modelo de Roy

Tobit vs. Truncamiento

Tobit: inversión en equipo

- $q_i^* = x_i\beta + \varepsilon_i$
- observamos $q_i = \begin{cases} q_i^* & \text{si } q_i^* > 0 \\ 0 & \text{si } q_i^* \leq 0 \end{cases}$
(algunas observaciones tienen inversión positiva, algunas observaciones no tienen inversión)

Truncamiento: Salarios

- $w_i = x_i\beta + \varepsilon_i$
- por razones de anonimidad, solo observamos (w_i, x_i) si $w_i \leq \bar{W}$
(el diseño del muestreo excluye salarios suficientemente grandes)

El modelo de selección de Heckman

- $w_i^* = x_i\beta + \varepsilon_i$
 - $s_i = \begin{cases} 1 & \text{if } \gamma'z_i + v_i > 0 \\ 0 & \text{if } \gamma'z_i + v_i \leq 0 \end{cases}$
 - observamos $w_i = w_i^*$ si $s_i = 1$
-
- la variable dependiente de interés, w_i^* , es *incidental* en el sentido de que depende de otra condición (la de participación)
 - si (ε, v) son conjuntamente normales, este es el modelo de selección de Heckman

El modelo de regresión normal con truncamiento

- $y = \beta_0 + \beta x + \varepsilon, \quad \varepsilon|x \sim N(0, \sigma^2)$
- solo observamos (y_i, x_i) si $y_i > 0$ (la muestra no es iid)
- en el modelo con truncamiento, solamente tenemos observaciones de una submuestra de la población seleccionada mediante el valor de la variable dependiente

¿Cuándo es MCO consistente?

- considera una regla de truncamiento genérica $s \in \{0, 1\}$ tal que $sy = \beta sx + s\varepsilon$
- puesto que $s^2 = s$, $E[(sx)(s\varepsilon)] = E[sx\varepsilon]$

OLS es inconsistente cuando $E[sx\varepsilon] \neq 0$
($E[s\varepsilon|x] \neq 0 \Rightarrow E[sx\varepsilon] \neq 0$)

- cuando s es independiente de ε , entonces OLS es consistente (incluso si s depende de x)
- en el modelo de truncamiento, $s = 1(\beta x + \varepsilon > 0)$, por lo que $E[s\varepsilon|x] \neq 0$ y MCO es inconsistente

Estimación MV

- La densidad de la muestra no es una normal porque la población ha sido truncada
- Necesitamos la distribución de y_i dado x_i Y dado que $y_i > 0$
- La densidad conjunta para $(y_i, y_i > 0)$ dado x_i : $\left(\frac{1}{\sigma}\right) \phi\left(\frac{\varepsilon_i}{\sigma}\right)$
- $Pr(y_i > 0|x_i) = \Phi\left(\frac{\beta x_i}{\sigma}\right)$

$$L_i(\beta, \sigma) = \frac{\left(\frac{1}{\sigma}\right) \phi\left(\frac{(y_i - \beta x_i)}{\sigma}\right)}{\Phi\left(\frac{\beta x_i}{\sigma}\right)}$$

Modelo de selección de Heckman

observamos w_i si $s_i = 1$

- ecuación output: $w = \beta_0 + \beta x + \varepsilon$
- ecuación de participación: $s = 1(\gamma'z + v)$
- $$\begin{bmatrix} \varepsilon \\ v \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\varepsilon^2 & \rho \\ \rho & 1 \end{bmatrix} \right)$$
- este modelo se puede generalizar para incluir otra ecuación output para aquéllos individuos tales que $s = 0$

MCO es inconsistente

- observa que $sw^* = s\beta_0 + \beta sx + s\varepsilon$
- entonces $E[sx * s\varepsilon | x, z] = E[s\varepsilon | x, z]x$ ya que $s^2 = s$
- por lo tanto, MCO estará sesgado si $E[s\varepsilon | x, z] \neq 0$

MCO es inconsistente si $\rho \neq 0$

Incluyendo controles adicionales

- incluir z en la ecuación output no resuelve el problema
- MCO falla porque en la muestra de trabajadores, la esperanza condicionada no es cero
- intuitivamente, los trabajadores con mayor probabilidad de trabajar también tenderán a tener mayores “errores” en salarios

Estimación MV

- es posible estimar el modelo por MV
- la expresión de la verosimilitud es más complicada que en los modelos Probit y Tobit ya que hay que obtener la distribución conjunta de w y s
- gretl puede ejecutar la estimación de Heckman por MV
- en general, la verosimilitud no es globalmente cóncava, y podemos tener máximos locales
- Heckman propuso un método sencillo en dos etapas basado en la esperanza condicionada que resulta en estimaciones consistentes

La esperanza condicionada

- del modelo Tobit sabemos que

$$E[w | x, z, s = 1] = x\beta + \rho\lambda(z\gamma)$$

- donde $\lambda()$ es la inversa de la razón de Mills
- λ se puede interpretar como una variable omitida correlacionada con $s\epsilon$
- si $\rho = 0$, no hay ningún problema en el estimador MCO

Corrección muestral en dos etapas

Corrección muestral en dos etapas de Heckman

- Primera etapa: usando todas las observaciones, estimamos un modelo Probit de *work* sobre z y calculamos la inversa del ratio de Mills, $\hat{\lambda}_i = \frac{\hat{\phi}_i}{\Phi_i}$
- Segunda etapa: usando la muestra de salarios, ols *wage* $\times \hat{\lambda}$

$\hat{\beta}$ es consistente y asintóticamente normal

¿Por qué funciona este método?

- Los estimadores MV de la ecuación de participación son consistentes
- $\hat{\lambda}$ desplaza la esperanza condicionada de cada individuo de acuerdo con su probabilidad de trabajar debido a factores inobservados. Supón que $\rho > 0$:
 - un individuo de la muestra de salarios con bajo índice $z_i\gamma$ (alto λ_i) tiene mayor probabilidad de trabajar por factores inobservados y probablemente salarios superiores debido a factores no observados: λ_i debería ser grande
 - un individuo de la muestra de salarios con alto índice $z_i\gamma$ (bajo λ_i) tiene menor probabilidad de trabajar por factores inobservados y probablemente salarios superiores debido a factores no observados: λ_i debería ser pequeño

Algunos asuntos sobre el procedimiento

- Las desviaciones típicas utilizando o1s (incluso con la opción `--robust`) no son correctas
- es posible contrastar la selección muestral (también en MV): estadístico t para $\hat{\rho}$ en la segunda etapa
- si hay controles endógenos en x , basta con aplicar `ts1s` en la segunda etapa
- el método funciona mejor si $x \subset z$ (es decir, algunas variables solo aparecen en la ecuación de participación)

El supuesto de normalidad

- malas noticias: el procedimiento solo es válido asintóticamente si los errores son normales (casi nunca lo son)
- buenas noticias: es fácil modificarlo para hacerlo robusto a
 - errores no-normales
 - errores heteroscedásticos

- el modelo de Roy es un modelo econométrico de dos sectores con autoselección
 - es muy influyente en la economía, sobre todo en la economía del trabajo y en el enfoque estructural para la evaluación de políticas.
 - fue primero desarrollado por Roy (1951) en su análisis de los ingresos en dos sectores ocupacionales, en los que los individuos se auto-seleccionan en el sector con las mayores ganancias.
- los agentes racionales optimizar decisiones sobre participación en diferentes mercados— trabajo, educación, el matrimonio, la delincuencia, etc.
- la auto-selección implica que las comparaciones directas de los resultados entre los individuos no son válidas para inferir relaciones causales

Segregación ocupacional por género y la brecha salarial

- Muchas mujeres trabajan en ocupaciones "femeninas" (empleo estable, flexible, sin depreciación del capital humano después de interrupciones de carrera) mientras que muchos hombres trabajan en ocupaciones "masculinas" (horas extras, acumulación de capital humano específico de la empresa)
 - En promedio, las mujeres ganan menos que los hombres. Es la brecha salarial de género debida a la discriminación de género?

Emigración y habilidad no observable

- Los emigrantes optan por ir al país anfitrión (con un alta prima por habilidad), mientras que los no-migrantes optan por permanecer en el país de origen (alta igualdad salarial)
 - En promedio, los emigrantes ganan menos que los no-migrantes en el país de acogida. ¿Significa esto que vienen de la parte baja en la distribución de habilidad del país de origen?

El modelo de Roy

observamos w_i tanto si $s_i = 1$ como si $s_i = 0$

- ecuación output si $s_i = 1$: $w_{1i} = x_i' \beta_1 + \varepsilon_1$
- ecuación output si $s_i = 0$: $w_{0i} = x_i' \beta_0 + \varepsilon_0$
- ecuación elección de sector: $s_i = 1 (z_i' \gamma + v_i)$

- $$\begin{bmatrix} \varepsilon_1 \\ \varepsilon_0 \\ v \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{10} & \rho_1 \\ \sigma_{10} & \sigma_0^2 & \rho_0 \\ \rho_1 & \rho_0 & 1 \end{bmatrix} \right)$$

Estimación

- La estimación por MV da estimaciones consistentes.
 - Sin embargo, los paquetes estadísticos estándar no proporcionan ningún comando: hay que programar la verosimilitud
- El procedimiento de estimación en dos etapas de Heckman sigue siendo válido:
 - se aplica el procedimiento para los trabajadores que eligen el sector 1 y se estima β_1 y el parámetro asociado a λ_1
 - se aplica el procedimiento a los trabajadores en el sector 0 y se estima β_0 y el parámetro asociado a λ_0
 - Para estimar σ_1^2 , σ_0^2 y σ_{10} hay que estimar modelos de $\text{var}(w_1|s=1, x)$ y $\text{var}(w_0|s=0, x)$ (más allá del objetivo de este curso)

Resumen

- hay una variedad de formas para controlar por el problema de selección de la muestra
- el modelo de Heckman asume errores normales y puede ser estimado tanto por MV como por el procedimiento de dos etapas de Heckman
- en el procedimiento de dos etapas los errores estándar de las estimaciones deben obtenerse teniendo en cuenta las dos etapas
- el modelo de Roy es un modelo básico de evaluación estructural de políticas económicas