Econometría Básica

Capítulo 17: Variable Limitada

José Valderrama & Freddy Rojas
jtvalderrama@gmail.com & frojasca@gmail.com
Universidad Católica Santo Toribio de Mogrovejo

Septiembre de 2021

Contenido

Contenido

Lógica

- El método de MV estima μ bajo la siguiente lógica:
  1. Los datos fueron generados con $N(\mu, \sigma^2 = 1)$
  2. Con los datos disponibles ¿cuál es el valor de μ que hace más probable que (1) sea cierto?
- Notar que típicamente se conoce μ y la distribución, y con esos datos se generan los número seudoaleatorios.
- En este caso es al revés, primero conocemos los datos, y con estos buscamos cuál fue el μ que los pudo haber generado.
- Máxima verosimilitud = Máxima compatibilidad entre el modelo y los datos.

Contenido

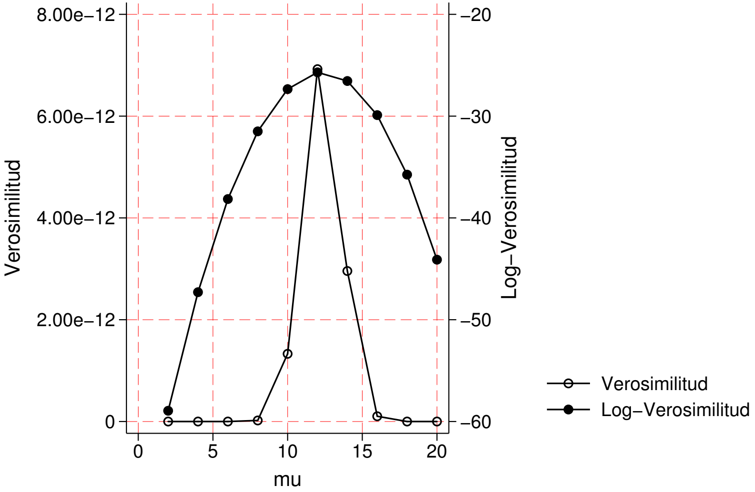
USAT Universidad Católica Santo Toribio de Mogrovejo

Máxima verosimilitud

- Dado un conjunto de datos, el objetivo es estimar los parámetros de tal manera que la muestra se parezca lo más posible al universo.
- El universo queda definido por la función de distribución que se asume tienen los datos (normal, exponencial, lognormal, etc)
- El logarítmo de la verosimilitud (log likelihood) es una transformación monotónica, por lo tanto...
- Mientras la verosimilitud ('l') ϵ $[0, 1]$ el log likelihood ('ll') ϵ...

USAT Universidad Católica Santo Toribio de Mogrovejo

Likelihood Vs Log-Likelihood



Estimación bajo MV

- Se trata de construir la función de probabilidad conjunta (o función de verosimilitud) de y_1, y_2,..., y_n suponiendo que las observaciones son independientes y están idénticamente distribuidas (iid)

$$L(\theta) = f(y_1 ... y_n; \theta) = \Pi_{i=1}^{n} f(y_i; \theta)$$

$$LL(\theta) = ln(f(y_1 ... y_n; \theta)) = \sum_{i=1}^{n} lnf(y_i; \theta)$$

- 'Si para un determinado valor de θ, la verosimilitud es *pequeña*, es poco probable que θ sea el valor correcto que ha generado los datos que observamos'

Estimación bajo MV

- Por tanto tenemos que elegir θ que maximice $L(\theta)$. Es decir, el estimador MV satisface la CPO:

$$\frac{\partial L(\theta)}{\partial \theta}|_{\theta=\hat{\theta}} = 0$$

- o lo que es lo mismo

$$\frac{\partial log(L(\theta))}{\partial \theta}|_{\theta=\hat{\theta}} = 0$$

- Y la condición de segundo orden...

Matrices relacionadas a la función Log(L)

- **Hessiana (H):** Es una matriz cuadrada (k x k) de las segundas derivadas de $log(L(\theta; y))$ con respecto a θ:

$$H(\theta) = \frac{\partial^2 Log(L(\theta; y))}{\partial\theta\partial\theta'}$$

- **Score (S):** Es una gradiente de (k x 1) de $log(L(\theta; y))$ con respecto a θ:

$$S(\theta) = \frac{\partial Log(L(\theta; y))}{\partial\theta}$$

Notar que cuando $\theta = \theta_{MV} \implies S(\theta_{MV}) = 0$

- **Matriz de información $I(\theta)$:** Indica el grado de curvatura

$$I(\theta) = E[-\frac{\partial^2 Log(L)}{\partial\theta\partial\theta'}]$$

Notar que mientras menos curvatura tenga la función de verosimilitud (el caso extremo es una línea recta) existirá mayor varianza en el estimador analizado pues:

$$Var(\theta) = [I(\theta)]^{-1}$$

USAT Universidad Católica Santo Toribio de Mogrovejo

Ventajas y desventajas

- **Ventaja:** El estimador MV (ML=maximum likelihood) tiene propiedades asintóticas óptimas entre todos los estimadores consistentes y normales asintóticamente.
- **Desventajas:**
  - El estimador ML depende de forma importante de los supuestos sobre la distribución.
  - El estimador MV tiene propiedades mediocres en muestras pequeñas.

Contenido

Pruebas asintóticas

- En econometría a menudo se plantean restricciones al modelo respecto a uno o más parámetros con el fin de indagar si el modelo es consistente con la restricción.
- Por ejemplo $\beta_0 = 1$ o $\alpha + \beta = 1$ en el contexto de una función de producción del tipo Cobb-Douglas (Retorno a escala constante).
- El modelo que se estima imponiendo la restricción precisamente se conoce como modelo restringido.

En lo que sigue se discuten tres pruebas asintóticas equivalentes que evalúan con procedimientos distintos la consistencia de una restricción.

Likelihood Ratio Test (LR)

Por sus siglas en inglés también se conoce como el "test LR". En este tipo de pruebas se requiere la estimación restringida y sin restringir:

- Modelo restringido: $y = \beta_0$
- Modelo no restringido: $y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2$
- ambos se estiman por MV donde lo que se quiere analizar es si la hipótesis nula conjunta de que los coeficientes de las variables que acompañan a las variables son iguales a cero...

$$LRT = 2[LogL(\hat{\theta}^n r) - LogL(\theta^r)] \sim \chi^2_q$$

Para su estimación requiere tanto de los estimadores restringidos como no restringidos.

USAT Universidad Católica Santo Toribio de Mogrovejo

Test de Wald

Piense en la siguiente restricción matricial: $R\beta = r$ donde R es una matriz de m x k y β es una matriz de k x 1. Sea $g(\beta) = R\beta - r$, si se sabe que $g(\hat{\beta}^R) = 0$, lo que se pregunta el test de Wald es ¿$g(\hat{\beta}_{MV}) = 0$? es decir, se reemplazan los $\hat{\beta}_{MV}$ en la restricción:

- Si $g(\hat{\beta}_{MV})$ tiende a 0 se acepta la restricción.
- Si $g(\hat{\beta}_{MV})$ no tiende a 0 no se acepta la restricción.

El estadístico necesario para la prueba es:

$$\begin{aligned} W &= (R\hat{\beta}_{MV} - r)'\{Var(R\hat{\beta}_{MV} - r)\}^{-1}(R\hat{\beta}_{MV} - r) \\ W &= g(\hat{\beta}_{MV})'\{\frac{\partial g}{\partial \beta}[I(\theta)]^{-1}\frac{\partial g}{\partial \beta}\}^{-1}g(\hat{\beta}_{MV}) \sim \chi^2_q \end{aligned}$$

Para su estimación requiere sólo de los estimadores no restringidos.

Test de Multiplicadores de Lagrange

Se basa en la matriz score eficiencia ($S(\theta)$), si se sabe que $S(\hat{\beta}_{MV}) = 0$, lo que se pregunta el test LM es ¿$S(\hat{\beta}^R) = 0$? es decir, se reemplazan los $\hat{\beta}^R$ en el score:

- Si $S(\hat{\beta}^R)$ tiende a 0 se acepta la restricción.
- Si $g(\hat{\beta}^R)$ no tiende a 0 no se acepta la restricción.

$$LM = S(\hat{\theta}^R)'[I(\theta)]^{-1}S(\hat{\theta}^R) \sim \chi^2_q$$

Para su estimación requiere sólo de los estimadores restringidos.

Contenido

MRS usando MV

Sea el siguiente modelo de regresión lineal simple, en su versión poblacional:

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i \tag{1}$$

Donde $\mu_i \sim N(0, \sigma^2)$, por lo tanto, la probabilidad de que el error μ_i provenga de la distribución normal es:

$$f(\mu_i, \beta_0, \beta_1, \sigma) = \frac{1}{\sqrt{\sigma^2 \pi}} \exp^{-\frac{1}{2}(\frac{\mu_i}{\sigma})^2} \tag{2}$$

Probabilidad conjunta

N° de Observación	f
1	$f(\mu_1, \beta_0, \beta_1, \sigma)$
2	$f(\mu_2, \beta_0, \beta_1, \sigma)$
.	.
.	.
.	.
n	$f(\mu_n, \beta_0, \beta_1, \sigma)$
Prob. Conjunta	$\pi_i^n f(\mu_i, \beta_0, \beta_1, \sigma)$

En una muestra de tamaño n la probabilidad individual de que cada observación provenga de una distribución normal son mostrados en la tabla 1. La probabilidad conjunta, asumiendo independencia, es la productoria de todas las probabilidades:

$$\pi_i^n f(\mu_i, \beta_0, \beta_1, \sigma) \quad = \quad L = \frac{1}{\sigma^n} \frac{1}{(\sqrt{2\pi})^n} \exp^{-\frac{1}{2\sigma^2} \sum_{i=1}^n \mu_i^2} \tag{3}$$

Con lo cual el logaritmo de la verosimilitud (LL) queda definido como:

$$LL \quad = \quad Ln(1) - Ln(\sigma)^n - Ln(\sqrt{2\pi})^n - \frac{1}{2\sigma^2} \sum \mu_i^2$$

Maximizar la expresión anterior es lo mismo que maximizar la siguiente función:

$$LL \quad = \quad -nLog(\sigma) - \frac{1}{2\sigma^2} \sum \mu_i^2 \tag{4}$$

La CPO de la función es dado por el siguiente arreglo:

$$S(\theta) = \begin{pmatrix} \frac{\partial LL}{\partial \beta_0} \\ \frac{\partial LL}{\partial \beta_1} \\ \frac{\partial LL}{\partial \sigma} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i) \\ \frac{1}{\sigma^2} \sum (Y_i - \beta_0 - \beta_1 X_i) X_i \\ \frac{-n}{\sigma} + \frac{1}{\sigma^3} \sum \mu_i^2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Nótese que los ecuaciones que se obtienen de las primeras dos filas de los vectores son las mismas ecuaciones normales que se obtienen cuando se resuelve el problema de MCO, por tanto los βs que resuelven el problema de MV son los mismos que los que se obtienen bajo MCO. La tercera fila del vector $S(\theta)$, permite conocer la dispersión de μ_i:

$$\sigma^2_{MV} = \frac{\sum \hat{\mu_i}^2}{n}$$

Que como se sabe es un estimador sesgado de la varianza[1], aunque el sesgo se disipa cuando la muestra (n) es grande.

La solución al problema todavía esta incompleto, falta demostrar que la Hessiana es una matriz definida negativa.

[1] El estimador insesgado es: $\frac{\sum \hat{\mu_i}^2}{n-k}$, donde k son todos los parámetros de la regresión a estimar incluyendo al intercepto

Los resultados anteriores pueden ser generalizados para el caso de más de un regresor o covariado, la estimación en este caso viene dado por:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}(X'Y) \\ \hat{\sigma}^2 &= n^{-1}\hat{\mu}'\hat{\mu} \end{aligned}$$

que se obtienen luego de maximizar la función log-likelihood:

$$LL(\theta; Y|X) = -\frac{n}{2}ln(2\pi) - \frac{n}{2}ln(\sigma^2) - \frac{1}{2\sigma^2}(Y - XB)'(Y - XB) \quad (5)$$

Comparación

MCO	MV
Función a optimizar:...........	Función a optimizar:...........
Criterio de optimización:...........	Criterio de optimización:...........
Rest. Vs No Rest:...........	Rest. Vs No Rest:...........
Estadístico:...........	Estadístico:...........
Interp. Est:...........	Interp. Est:...........
P-Value:...........	P-Value:...........
Bondad de ajuste:.............	Bondad de ajuste:...........

Contenido

Motivación

Al finalizar la sesión usted debe estar en la capacidad de:

- Definir los modelos de probabilidad lineal y no lineal.
- Exponer las principales características de cada modelo.
- Estimar modelos simples no lineales empleando la técnica de MV.
- Identificar las situaciones en las que es más apropiado emplear un tipo de modelo u otro.

Considere los siguientes problemas en economía:

- Estimar la probabilidad de que un nuevo cliente sea buen o mal pagador.
- Estimar un modelo de oferta laboral cuya dependiente es la dicotómica participa o no en la PEA.
- En general, estimar modelos donde la variable dependiente es binaria o dicotómica.

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

El modelo de probabilidad lineal (MPL)

- La variable dependiente es una dummy:
  - $y_i = 1$, si se cumple con el criterio.
  - $y_i = 0$, si no se cumple con el criterio.
- El modelo es: $y = x\beta + \epsilon$
- Cuando y es una variable binaria, entonces: $P(y_i = 1) = E[y_i] = x\beta$
- Si estimáramos este modelo con un Modelo de Probabilidad Lineal (es decir una regresión MCO):
  - Predicciones con valores fuera del rango $[0, 1]$
  - Heterocedasticidad: $Var[\epsilon|x] = x'\beta(1 - x'\beta)$
  - Distribución no normal de la perturbación aleatoria

USAT Universidad Católica Santo Toribio de Mogrovejo

Modelos no lineales

- Se asume existencia de una variable latente y^*
- y^* determina el valor de y (lo observable)
  - $y_i = 1$ si sólo sí: $y_i^* = X_i\beta + \varepsilon > 0$
  - $y_i = 0$ si sólo sí: otro caso
- Entonces vamos a observar $y = 1$ sólo cuando: $\varepsilon_i > -X_i\beta$
- Siendo F la función de densidad acumulada de la variable aleatoria ε_i entonces la probabilidad que $y = 1$ es:

$$P(y_i = 1) = P(\varepsilon_i > -X_i\beta) = 1 - F(-X_i\beta) = F(X_i\beta)$$

- La forma de F dependerá de la distribución de ε_i
  - Si se asume que ε_i se distribuye según una función normal: *Modelo Probit*
  - Si se asume que ε_i se distribuye según una función logística: *Modelo Logit*

La crítica al MPL acerca de las predicciones fuera de rango pueden ser contestadas con modelos no lineales (Ver gráfico)

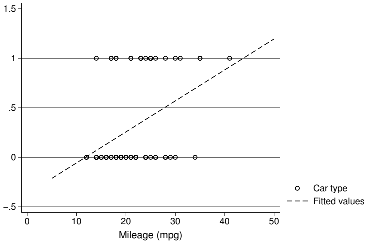


FIGURA: Modelo de Probabilidad Lineal (MPL)

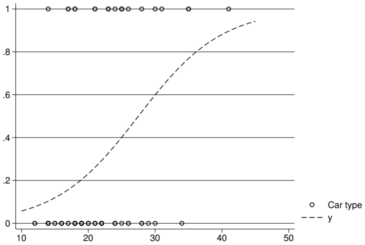


FIGURA: Modelo de Probabilidad No lineal

En teoría son varias las funciones no lineales que se pueden usar pero las mas populares son:

MODELO PROBIT $F(\omega) = \int_{-\infty}^{\omega} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$

MODELO LOGIT $F(\omega) = L(\omega) = \frac{e^{\omega}}{1+e^{\omega}}$

Fundamento económico

Al menos dos justificaciones

1. Precio de reserva
2. Modelo de utilidad aleatoria

Efecto Marginal

El efecto marginal para la persona i en la variable k, cuando la variable x_k es continua se define como:

$$EM_{ik} = \frac{\partial Prob(Y_i = 1)}{\partial x_k} = \frac{\partial F(X_i'\beta)}{\partial x_k} = \beta_k f(X_i'\beta)$$

Mientras que cuando esta es dicotómica el efecto se define como:

$$EM_{ik} = F(X_i'\beta | x_k = 1) - F(X_i'\beta | x_k = 0)$$

USAT Universidad Católica Santo Toribio de Mogrovejo

Comparación de Efectos Marginales

Una forma de comparar los resultados de los diferentes modelos es empleando efectos marginales:

$$\frac{\partial}{\partial x_{ik}} x_i' \beta = \beta_k^{PL} \tag{6}$$

$$\frac{\partial}{\partial x_{ik}} \Phi(x_i'\beta) = \phi(x_i'\beta) \cdot \beta_k^{Probit} \tag{7}$$

$$\frac{\partial}{\partial x_{ik}} L(x_i'\beta) = \frac{e^{x_i'\beta}}{(1+e^{x_i'\beta})^2} \cdot \beta_k^{Logit} \tag{8}$$

Así, cuando $x_i'\beta = 0$, se pueden tener "equivalencias" entre los distintos βs al igualar los efectos marginales de (1) con (2) y los efectos marginales de (1) y (3).

USAT Universidad Católica Santo Toribio de Mogrovejo

Odds Ratios

En el caso de un logit del tipo:

$$Pr[Y_j = 1|X_j] = \frac{exp(\alpha_0 + \beta_0 X_j)}{1 + exp(\alpha_0 + \beta_0 X_j)} \quad (9)$$

Los odds se definen como:

$$Odds(X) = \frac{Pr[Y_j = 1|X_j]}{Pr[Y_j = 0|X_j]} = \frac{F(\alpha_0 + \beta_0 X_j)}{1 - F(\alpha_0 + \beta_0 X_j)} = exp(\alpha_0 + \beta_0 X_j) \quad (10)$$

Los odds ratios es el ratio de dos odds para diferentes valores de X_j, digamos $X_j = x$ y $X_j = x + \Delta x$:

$$\frac{Odds(x + \Delta x)}{Odds(x)} = \frac{exp(\alpha + \beta x + \beta \Delta x)}{exp(\alpha + \beta x)} = exp(\beta \Delta x)$$

Contenido

ESTIMACIÓN

- Se emplea MV
- Sea una muestra $y_1, y_2..., y_n$ cuya probabilidad de ocurrir es $P(muestra)$
- $P(muestra) = P(y_1) * P(y_2) * P(y_2)... * P(y_n)$, si cada y_i es independiente
- Donde $P(y_i) = 1 - F(X_i\beta)$ si $y1 = 1$ o $P(y_i) = F(X_i\beta)$ si $y_i = 0$
- La función de verosimilitud:

$$L = \Pi_{y_i=1}[1 - F(-X_i\beta)] * \Pi_{y_i=0}F(-X_i\beta)$$

- Los βs los encontramos maximizando esta función (matemáticamente es un proceso iterativo, se empieza con un grupo de βs y se va iterando hasta alcanzar los βs óptimos)

Modelo ingenuo

Estime el modelo logit ingenuo sabiendo que:

Y
1
0
0

Generalización

Sea (y_i, x_i) con $i = 1, ..., n$, una muestra *iid* Y_i tiene una distribución de Bernoulli con $p_i = Pr(y_i = 1)$ con lo cual la función de verosimilitud será:

$$L(\beta) = \prod_{y_i=1} p_i \prod_{y_i=0} (1 - p_i) = \prod_{i=1}^{n} p_i^{y_i} (1 - p_i)^{1-y_i}$$

y su logaritmo:

$$\begin{aligned} l(\beta) &= \sum_{i=1}^{n} [y_i(lnp_i) + (1 - y_i)ln(1 - p_i)] \\ &= \sum_{i=1}^{n} [y_i lnF(x_i'\beta) + (1 - y_i)ln(1 - F(x_i'\beta))] \end{aligned}$$

USAT Universidad Católica Santo Toribio de Mogrovejo

Generalización

Siendo las condiciones de primer orden:

$$\sum_{i=1}^{n} \frac{(y_i - F_i) f_i}{F_i(1 - F_i)} x_{ki} \quad = \quad 0$$

con $k = 1, ..., K$; $F_i \equiv F(x_i'\beta)$ y $f_i \equiv f(x_i'\beta)$

En el caso particular de un logit ($F(x) = \frac{exp(x)}{1+exp(x)}$) la expresión se reduce a:

$$\frac{\partial \ln L}{\partial \theta} \quad = \quad \sum_{i=1}^{n} (y_i - F_i)\mathbf{x}_i = \mathbf{0}$$

USAT Universidad Católica Santo Toribio de Mogrovejo

Contenido

Curvas ROC

- ¿Cómo predecir la ocurrencia de un evento a partir de la estimación de un modelo tipo probit o logit? (buen o mal pagador)
- Los modelos permiten calcular las probabilidades de ocurrencia, típicamente se asume que si la probabilidad predicha es superior a 0.5, entonces se asume que el evento ocurrirá.
- Reconocer que existen dos tipos de errores que se cometen cuando se hace la predicción discreta es la motivación de la estrategia ROC.
- El porcentaje de unos clasificados incorrectamente y el porcentaje de ceros clasificados incorrectamente se conocen como sensitividad y especificidad, respectivamente.

USAT Universidad Católica Santo Toribio de Mogrovejo

Curvas ROC

	Verdadero		
Clasificado	D=1	D=0	Total
+	a	c	a+c
-	b	d	b+d
Total	a+b	c+d	a+b+c+d

Cuadro: Classified es + si Pr(D) predicha >= umbral

Donde "Verdadero" hace referencia a lo efectivamente observado (lo verdadero), D=1 y D=0 y "Classified" es lo estimado probabilísticamente.

Curvas ROC

Si la probabilidad de ocurrencia es mayor a un determinado umbral (por ejemplo 0.5), entonces se clasifica como cierto la ocurrencia (+). A partir de lo predicho y lo efectivamente observado se pueden construir los siguientes indicadores:

1. `Sensitibidad:` Porcentaje de unos clasificados correctamente: $\frac{a}{a+b}$
2. `Especificidad:` Porcentaje de ceros clasificados correctamente: $\frac{d}{c+d}$
3. `Correctamente clasificados:` Total de correctamente clasificados: $\frac{a+d}{a+b+c+d}$

Curvas ROC. Presentación gráfica

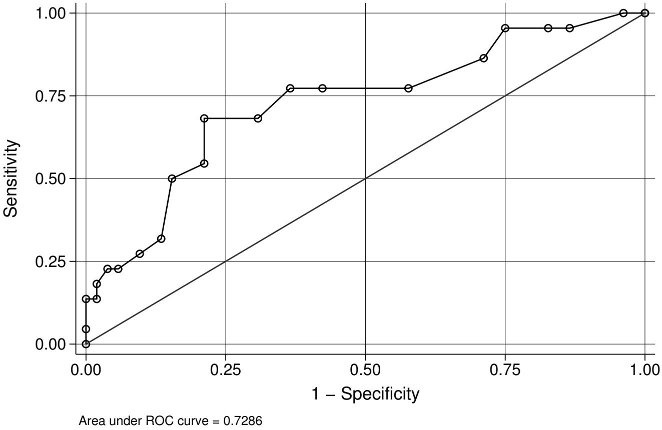


Figura: Curva ROC.

Curvas ROC. Selección de modelos

Bondad de ajuste

- Pseudo-R2 que compara la función de verosimilitud maximizada por nuestros betas con una función donde todos los betas son cero y solo hay una constante (modelo ingenuo)
- Formalmente: Test LR (Likelihood ratio)

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

Contenido

Motivación

1. Ejemplo 1: En una encuesta de hogares se observa el salario solo de aquellos que trabajan, el resto de sus características son conocidas (sexo, edad, etc)
2. Ejemplo 2: Se encuesta a mujeres que trabajan.
3. Ejemplo 3: Los resultados de una encuesta televisada son representativos de todos los votantes?.

Contenido

Propiedades distribución normal

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
\phi(z) &= \frac{1}{\sqrt{2\pi}} exp\left[-\frac{(z)^2}{2}\right]
\end{aligned}
$$

- $\phi(-z) = \phi(z)$
- $\frac{\partial \phi(z)}{\partial z} = -z\phi(z)$
- $f(x) = \frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right] = \frac{1}{\sigma}\phi(z)$
- $\Phi(a) = P(z < a) = \int_{-\infty}^{a} \phi(z)dz$
- $\Phi(-a) = 1 - \Phi(a) = P(z \geq a)$

USAT Universidad Católica Santo Toribio de Mogrovejo

Propiedades distribución normal

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$
$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{(z)^2}{2}\right]$$

- $\phi(-z) = \phi(z)$
- $\frac{\partial\phi(z)}{\partial z} = -z\phi(z)$
- $f(x) = \frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right] = \frac{1}{\sigma}\phi(z)$
- $\Phi(a) = P(z < a) = \int_{-\infty}^{a}\phi(z)dz$
- $\Phi(-a) = 1 - \Phi(a) = P(z \geq a)$

Propiedades distribución normal

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
\phi(z) &= \frac{1}{\sqrt{2\pi}} exp\left[-\frac{(z)^2}{2}\right]
\end{aligned}
$$

- $\phi(-z) = \phi(z)$
- $\frac{\partial\phi(z)}{\partial z} = -z\phi(z)$
- $f(x) = \frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right] = \frac{1}{\sigma}\phi(z)$
- $\Phi(a) = P(z < a) = \int_{-\infty}^{a} \phi(z)dz$
- $\Phi(-a) = 1 - \Phi(a) = P(z \geq a)$

Propiedades distribución normal

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$\phi(z) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{(z)^2}{2}\right]$$

- $\phi(-z) = \phi(z)$
- $\frac{\partial\phi(z)}{\partial z} = -z\phi(z)$
- $f(x) = \frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right] = \frac{1}{\sigma}\phi(z)$
- $\Phi(a) = P(z < a) = \int_{-\infty}^{a} \phi(z)dz$
- $\Phi(-a) = 1 - \Phi(a) = P(z \geq a)$

Propiedades distribución normal

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
\phi(z) &= \frac{1}{\sqrt{2\pi}} exp\left[-\frac{(z)^2}{2}\right]
\end{aligned}
$$

- $\phi(-z) = \phi(z)$
- $\frac{\partial \phi(z)}{\partial z} = -z\phi(z)$
- $f(x) = \frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right] = \frac{1}{\sigma}\phi(z)$
- $\Phi(a) = P(z < a) = \int_{-\infty}^{a} \phi(z)dz$
- $\Phi(-a) = 1 - \Phi(a) = P(z \geq a)$

Propiedades distribución normal

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \\
\phi(z) &= \frac{1}{\sqrt{2\pi}} exp\left[-\frac{(z)^2}{2}\right]
\end{aligned}
$$

- $\phi(-z) = \phi(z)$
- $\frac{\partial\phi(z)}{\partial z} = -z\phi(z)$
- $f(x) = \frac{1}{\sigma}\phi\left[\frac{x-\mu}{\sigma}\right] = \frac{1}{\sigma}\phi(z)$
- $\Phi(a) = P(z < a) = \int_{-\infty}^{a} \phi(z)dz$
- $\Phi(-a) = 1 - \Phi(a) = P(z \geq a)$

USAT Universidad Católica Santo Toribio de Mogrovejo

Distribución truncada

- $f(x|x > a) = \frac{f(x)}{P(x>a)}$
- $E(x|x > a) > E(x)$
- $Var(x|x > a) < Var(x)$

Universidad Católica Santo Toribio de Mogrovejo

Distribución normal truncada

Sea x una variable aleatoria distribuida normalmente: $x \sim N(\mu, \sigma^2)$, entonces:

- $\lambda(\alpha) = \frac{\phi(\alpha)}{1-\Phi(\alpha)}$ donde $\alpha = \frac{a-\mu}{\sigma}$. Evaluar y graficar en $\alpha = 0$
- $\delta(\alpha) = \lambda(\alpha)[\lambda(\alpha) - \alpha]$
- $0 < \delta(\alpha) < 1$
- $\lambda(\alpha)$ es conocido como el inverso del ratio de Mills.
- $E(x|x > a) = \mu + \sigma\lambda(\alpha)$
- $Var(x|x > a) = \sigma^2[1 - \delta(\alpha)]$

Contenido

Caso: Promedio truncado NO condicionado

Por medio de una encuesta a empresas formales sólo los trabajadores con ingresos superiores a S/.930 son declarados. Si se estima que el porcentaje de trabajadores con un sueldo superior a S/.930 es el 80 % de los trabajadores, y que el salario promedio de los entrevistados es de S/.1,800 Nuevos Soles. Estime el salario promedio de todos los trabajadores.

Solución:

$$E(x|x > 930) = \mu + \sigma \frac{\phi[(930-\mu)/\sigma]}{1-\Phi} \quad (11)$$

$$\Phi(\alpha) = 20\,\%; \alpha = (930-\mu)/\sigma \quad (12)$$

Ayuda, utilice los comandos `-invnormal-` y `-normalden-`

Caso: Promedio truncado NO condicionado

Por medio de una encuesta a empresas formales sólo los trabajadores con ingresos superiores a S/.930 son declarados. Si se estima que el porcentaje de trabajadores con un sueldo superior a S/.930 es el 80 % de los trabajadores, y que el salario promedio de los entrevistados es de S/.1,800 Nuevos Soles. Estime el salario promedio de todos los trabajadores.

Solución:

$$E(x|x > 930) = \mu + \sigma \frac{\phi[(930 - \mu)/\sigma]}{1 - \Phi} \tag{11}$$

$$\Phi(\alpha) = 20\,\%; \alpha = (930 - \mu)/\sigma \tag{12}$$

Ayuda, utilice los comandos `-invnormal-` y `-normalden-`

El modelo de regresión truncado

La muestra está sistemáticamente restringida a solo una parte de la población. Por ejemplo, una muestra puede solo incluir personas que están empleadas, o gente sobre una cierta edad.

$Y_i = X_i\beta + \mu_i;\ \mu_i \sim N(0, \sigma^2)$

$Y_i|X_i \sim N[X_i\beta, \sigma^2]$

$$E[y_i|y_i > a] = X_i\beta + \sigma\frac{\phi(\alpha_i)}{1 - \Phi(\alpha_i)}$$

donde:

$\alpha_i = \frac{a - X_i\beta}{\sigma}$ y ϕ y Φ son las funciones de densidad y acumulada de una variable aleatoria normal estándar.

$$E[y_i|y_i > a] = X_i\beta + \sigma\lambda(\alpha_i)$$

El modelo de regresión truncado

La muestra está sistemáticamente restringida a solo una parte de la población. Por ejemplo, una muestra puede solo incluir personas que están empleadas, o gente sobre una cierta edad.

$Y_i = X_i\beta + \mu_i;\ \mu_i \sim N(0, \sigma^2)$

$Y_i | X_i \sim N[X_i\beta, \sigma^2]$

$$E[y_i | y_i > a] = X_i\beta + \sigma \frac{\phi(\alpha_i)}{1 - \Phi(\alpha_i)}$$

donde:

$\alpha_i = \frac{a - X_i\beta}{\sigma}$ y ϕ y Φ son las funciones de densidad y acumulada de una variable aleatoria normal estándar.

$$E[y_i | y_i > a] = X_i\beta + \sigma\lambda(\alpha_i)$$

El modelo de regresión truncado

La muestra está sistemáticamente restringida a solo una parte de la población. Por ejemplo, una muestra puede solo incluir personas que están empleadas, o gente sobre una cierta edad.

$Y_i = X_i\beta + \mu_i;\ \mu_i \sim N(0, \sigma^2)$

$Y_i|X_i \sim N[X_i\beta, \sigma^2]$

$$E[y_i|y_i > a] = X_i\beta + \sigma\frac{\phi(\alpha_i)}{1-\Phi(\alpha_i)}$$

donde:

$\alpha_i = \frac{a-X_i\beta}{\sigma}$ y ϕ y Φ son las funciones de densidad y acumulada de una variable aleatoria normal estándar.

$$E[y_i|y_i > a] = X_i\beta + \sigma\lambda(\alpha_i)$$

El modelo de regresión truncado

La muestra está sistemáticamente restringida a solo una parte de la población. Por ejemplo, una muestra puede solo incluir personas que están empleadas, o gente sobre una cierta edad.

$Y_i = X_i\beta + \mu_i;\ \mu_i \sim N(0, \sigma^2)$

$Y_i|X_i \sim N[X_i\beta, \sigma^2]$

$$E[y_i|y_i > a] = X_i\beta + \sigma\frac{\phi(\alpha_i)}{1-\Phi(\alpha_i)}$$

donde:

$\alpha_i = \frac{a-X_i\beta}{\sigma}$ y ϕ y Φ son las funciones de densidad y acumulada de una variable aleatoria normal estándar.

$$E[y_i|y_i > a] = X_i\beta + \sigma\lambda(\alpha_i)$$

El modelo de regresión truncado

La muestra está sistemáticamente restringida a solo una parte de la población. Por ejemplo, una muestra puede solo incluir personas que están empleadas, o gente sobre una cierta edad.

$Y_i = X_i\beta + \mu_i$; $\mu_i \sim N(0, \sigma^2)$

$Y_i|X_i \sim N[X_i\beta, \sigma^2]$

$$E[y_i|y_i > a] = X_i\beta + \sigma\frac{\phi(\alpha_i)}{1 - \Phi(\alpha_i)}$$

donde:

$\alpha_i = \frac{a - X_i\beta}{\sigma}$ y ϕ y Φ son las funciones de densidad y acumulada de una variable aleatoria normal estándar.

$$E[y_i|y_i > a] = X_i\beta + \sigma\lambda(\alpha_i)$$

El modelo de regresión truncado

La muestra está sistemáticamente restringida a solo una parte de la población. Por ejemplo, una muestra puede solo incluir personas que están empleadas, o gente sobre una cierta edad.

$Y_i = X_i\beta + \mu_i;\ \mu_i \sim N(0, \sigma^2)$

$Y_i|X_i \sim N[X_i\beta, \sigma^2]$

$$E[y_i|y_i > a] = X_i\beta + \sigma\frac{\phi(\alpha_i)}{1 - \Phi(\alpha_i)}$$

donde:

$\alpha_i = \frac{a - X_i\beta}{\sigma}$ y ϕ y Φ son las funciones de densidad y acumulada de una variable aleatoria normal estándar.

$$E[y_i|y_i > a] = X_i\beta + \sigma\lambda(\alpha_i)$$

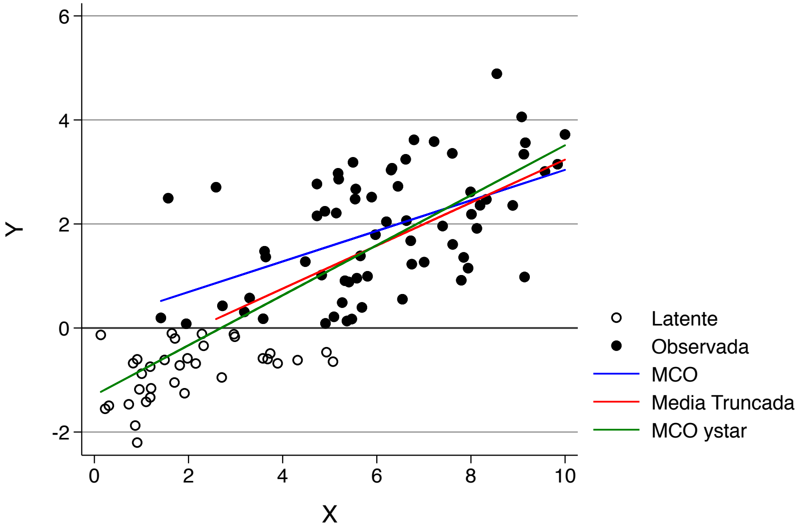
El modelo a estimar sería entonces:

$$y_i = X_i\beta + \sigma\lambda(\alpha_i) + \varepsilon_i$$

Donde $\alpha_i = \frac{a - X_i\beta}{\sigma}$

Así, si se estima el modelo sin incorporar λ se tendría un problema de variable omitida los parámetros serían sesgados e inconsistentes.

Regresión truncada versus MCO



Efectos marginales

- La interpretación de los parámetros dependen de la pregunta de investigación.
- Si el investigador esta interesado en la relación entre las variables para la población entera, los coeficientes β son interpretados como los efectos marginales.
- Sin embargo, si el interés en el efecto en la submuestra, el efecto marginal se estima de la siguiente manera:

$$\frac{\partial E[y_i|y_1 > a]}{\partial X} = \beta + \sigma \frac{\partial \lambda}{\partial \alpha}\frac{\partial \alpha}{\partial X} = \beta(1 - \delta(\alpha))$$

- Es decir, el efecto marginal es inferior que β, pues $0 < \delta(\alpha) < 1$.

Estimación

$$ln\ell = \sum_{i=1}^{N} ln\left[\sigma^{-1}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] - \sum_{i=1}^{N} ln\left[1 - \Phi\left(\frac{a - x_i'\beta}{\sigma}\right)\right]$$

Contenido

El modelo de regresión censurado

- Un modelo es censurado cuando los valores de la variable dependiente se encuentran restringidos a un rango de valores.
- La diferencia con el truncamiento es que mientras en este modelo sólo se observa una submuestra, en censura se tiene información de las variables independientes de toda la muestra.
- Dos ejemplos famosos por el uso de esta metodología son los realizados por Tobin (1958) y por Fair (1978).
- El primero estudió los determinantes del gasto en bienes durables (dichos gastos se acumulan en 0 y luego tienen valores positivos), mientras que el segundo estudió el número de relaciones extramaritales[2] .

[2] Aunque Fair usó un modelo tipo Tobit se puede argumentar que un modelo tipo *"Count Data"* es más apropiado.

El modelo (Tobit tipo 1)

Considere un modelo estándar probit basado en la decisión de comprar algo. Sea y^* una variable índice o latente que se asume puede ser modelado como:

$$y_i^* = x_i'\beta + \epsilon_i \text{ donde } \epsilon_i \sim N(0, \sigma^2)$$

Lo que determina:

- $y_i = 1$ si $y_i^* > 0$
- $y_i = 0$ si $y_i^* <= 0$

Si además de conocer si el individuo compra o no, se conoce lo gastado en la compra, se tendría el modelo conocido como Tobit[3]:

- $y_i = y_i^*$ si $y_i^* > 0$
- $y_i = 0$ si $y_i^* <= 0$

[3] El nombre es en honor de Tobin, quién en 1958 fue el primero que lo propuso

El modelo (Tobit tipo 1)

Resumen de lo anterior es:

$$y_i = max(0, x_i'\beta)$$

$$\begin{aligned}
E(y_i|x_i) &= 0 \cdot P(y_i^* \leq 0|x_i) + E(y_i^*|y_i^* > 0, x_i) \cdot P(y_i^* > 0|x_i) \\
&= E(y_i^*|y_i^* > 0, x_i) \cdot P(y_i^* > 0|x_i) \\
&= (\mu + \sigma\lambda) \cdot P(x_i'\beta + \epsilon_i > 0|x_i) \\
&= (x_i'\beta + \sigma\frac{\phi(x_i'\beta/\sigma)}{\Phi(x_i'\beta/\sigma)}) \cdot P(\epsilon_i/\sigma > -x_i'\beta/\sigma|x_i) \\
&= (x_i'\beta + \sigma\frac{\phi(x_i'\beta/\sigma)}{\Phi(x_i'\beta/\sigma)}) \cdot \Phi(x_i'\beta/\sigma) \\
&= x_i'\beta\Phi(x_i'\beta/\sigma) + \sigma\phi(x_i'\beta/\sigma)
\end{aligned}$$

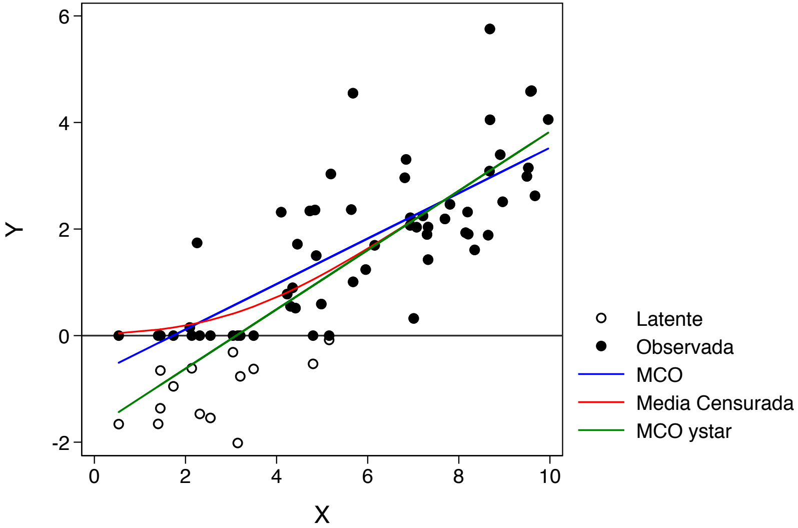
USAT Universidad Católica Santo Toribio de Mogrovejo

Interpretación de los parámetros

- Dependen si estamos interesados en saber algo sobre la media en la distribución censurada o los coeficientes del modelo latente.
- $\frac{\partial E(y_i|y_i>0,x_i)}{\partial x_j} = \beta_j \left[1 - \lambda\left(\frac{x_i\beta}{\sigma}\right)\left[\frac{x_i\beta}{\sigma} + \lambda\left(\frac{x_i\beta}{\sigma}\right)\right]\right] < \beta_j$
- $\frac{\partial E(y_i|x_i)}{\partial x_j} = \beta_j \Phi(\frac{x_i\beta}{\sigma}) < \beta_j$
- $\frac{\partial E(y_i^*|x)}{\partial x_j} = \beta_j$

USAT Universidad Católica Santo Toribio de Mogrovejo

El modelo (Tobit tipo 1)



Estimación

- La estimación MCO de la variable observada y_i sobre x_i $y_i = x_i'\beta + \mu_i$ produce estimadores sesgados de β, pues $E(y_i|x_i) = x_i'\beta\Phi(x_i'\beta/\sigma) + \sigma\phi(x_i'\beta/\sigma)$ no es una función lineal de x_i.
- Incluso restringir la muestra a las observaciones que son observadas $(y_i > 0)$ no resuelven el problema, pues como se sabe la especificación de la regresión truncada es $E(y_i|y_i > 0, x_i) = x_i'\beta + \sigma\lambda(x_i'\beta/\sigma)$, es decir, de estimar el modelo sin incluir λ se tendría un problema de variable omitida.

Estimación

- Asumiendo independencia entre observaciones, el modelo Tobit es usualmente estimado por MV, siendo necesario entonces definir la función de verosimilitud.
- A diferencia de la función de densidad de una distribución truncada, que es escalada por el término $1/(1-\Phi(\alpha))$ para que la masa de probabilidad siga siendo igual a 1, en el caso censurado la función de densidad es una combinación de dos funciones, una discreta y otra continua.
- La discreta es donde se acumulan los puntos luego del truncamiento (0 por ejemplo) y la continua es la función de densidad convencional.

Estimación

$$
\begin{aligned}
Prob(y_i = 0) &= Prob(y_i^* \leq 0) = Prob(-x_i'\beta \geq \epsilon_i) \\
&= Prob(-x_i'\beta/\sigma \geq \epsilon_i/\sigma) \\
&= 1 - \Phi(\frac{x_i'\beta}{\sigma}) \\
Prob(y_i = y_i^*) &= Prob(y_i = x_i'\beta + \epsilon_i) = Prob(\epsilon_i = y_i - x_i'\beta) \\
&= f(y_i - x_i'\beta) \\
&= \frac{1}{\sigma}\phi((y_i - x_i'\beta)/\sigma)
\end{aligned}
$$

Estimación

Con lo cual, la función de verosimilitud queda definido como:

$$\ell \quad = \quad \Pi_{y_i=0}\left[1-\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]\cdot\Pi_{y_i=y_i^*}\left[\frac{1}{\sigma}\phi\left(\frac{y_i-x_i'\beta}{\sigma}\right)\right]$$

y la función log-l:

$$ln\ell \quad = \quad \sum_{y_i=0}\left[1-\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]+\sum_{y_i=y_i^*}\left[-\frac{1}{2}ln(2\pi)+ln(\sigma^2)+\frac{(y_i-x_i'\beta)^2}{\sigma^2}\right]$$

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

Modelo de Heckman

$$
\begin{aligned}
y_{1i} &= x_{1i}\beta_1 + \mu_{1i} \quad \text{(Ecuación de regresión)} && (13)\\
y_{2i}^* &= x_{2i}\beta_2 + \mu_{2i} \quad \text{(Ecuación de selección)} && (14)\\
y_{2i} &= 1[y_{2i}^* > 0] \quad \text{(Dicotómica de selección)} && (15)
\end{aligned}
$$

Modelo de Heckman

Por ejemplo:

1. Regresión: Ecuación de salarios, donde x_1 son los determinantes de la productividad.
2. Selección: Decisión de trabajar, donde y_2^* puede ser leído como la utilidad neta del trabajo y x_2 son los determinantes de la decisión de trabajar.
3. Dicotómica que clasifica a las personas según estas trabajen o no (ocurre cuando $y_2^* > 0$)

USAT Universidad Católica Santo Toribio de Mogrovejo

Contenido

Modelo de Heckman

1. (y_{2i}, x_{2i}) información disponible para todos.
2. (y_{1i}, x_{1i}) información disponible solo si $y_{2i} = 1$ (muestra bajo selección).
3. (μ_{1i}, μ_{2i}) independientes de x_{2i}, con esperanzas nulas.
4. $\mu_{2i} \sim N(0, \sigma_2^2)$
5. $E(\mu_{1i}|\mu_{2i}) = \gamma\mu_{2i}$

Donde la ecuación (5) es la que determina que el modelo sea de selección y no una regresión ordinaria.

Contenido

Modelo de Heckman: Problema

$$
\begin{aligned}
E(y_{1i}|x_{1i}, y_{2i} = 1) &= x_{1i}\beta_1 + E(\mu_{1i}|x_{1i}, y_{2i} = 1) && (16)\\
&= x_{1i}\beta_1 + E(E(\mu_{1i}|\mu_{2i})|x_{1i}, y_{2i} = 1) && (17)\\
&= x_{1i}\beta_1 + E(\gamma\mu_{2i}|x_{1i}, y_{2i} = 1) && (18)\\
&= x_{1i}\beta_1 + \gamma E(\mu_{2i}|x_{1i}, y_{2i}^* > 0) && (19)\\
&= x_{1i}\beta_1 + \gamma E(\mu_{2i}|x_{1i}, \mu_{2i}^* > -x_{2i}\beta_2) && (20)\\
&= x_{1i}\beta_1 + \gamma[E(\mu_{2i} + \sigma_2\lambda(-x_{2i}\beta_2/\sigma_2)] && (21)\\
&= x_{1i}\beta_1 + \gamma\sigma_2[\lambda(-x_{2i}\beta_2/\sigma_2)] && (22)\\
&= x_{1i}\beta_1 + \gamma\sigma_2 z_i \neq x_{1i}\beta_1 && (23)
\end{aligned}
$$

Problema: Estimar MCO con la muestra bajo selección es inconsistente.

USAT Universidad Católica Santo Toribio de Mogrovejo

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

Modelo de Heckman: Solución 1/4

- Heckman (1979) estudió el sesgo de selección como un problema de especificación incorrecta.
- En este caso la variable omitida z_i genera inconsistencia.
- Donde la fuente de la inconsistencia es la correlación entre μ_{1i} y μ_{2i}.
- En particular que en $E(\mu_{1i}|\mu_{2i}) = \gamma\mu_{2i}$, $\gamma \neq 0$
- La gran contribución de Heckman fue notar que el sesgo de selección es un sesgo por omisión de variables.
- La fuente de la selección es la correlación entre el mecanismo de selección y la regresión a través de sus errores.
- Usando LEI, se puede probar que $E(\mu_1, \mu_2) = \gamma\sigma_2^2$

Modelo de Heckman: Solución 2/4

$$E(y_{1i}|x_{1i}, y_{2i} = 1) = x_{1i}\beta_1 + \gamma\sigma_2 z_i \quad (24)$$

$$y_{1i} = x_{1i}\beta_1 + \gamma\sigma_2 z_i + \mu^*_{1i} \quad (25)$$

- Si x_{1i} , z_i fuesen observables cuando $y_{2i} = 1$; se puede estimar (13) por MCO.
- El problema es que $z_i = \lambda(-x_{2i}\beta_2/\sigma_2)$ no es observable, pues β_2 y σ_2 no son conocidos.
- No es necesario conocer los valores de dichos parámetros por separado, con estimar $\delta = \beta_2/\sigma_2$ es suficiente como para tener un estimador de z_i.

USAT Universidad Católica Santo Toribio de Mogrovejo

Modelo de Heckman: Solución 3/4

Se puede estimar δ usando un probit:

$$\begin{aligned}
P(y_{2i}=1) &= P(y_{2i}^{*}>0) = P(x_{2i}\beta_2+\mu_{2i}>0) && (26)\\
&= P(\mu_{2i}>-x_{2i}\beta_2) && (27)\\
&= P(\frac{\mu_{2i}}{\sigma_2}>-\frac{x_{2i}\beta_2}{\sigma_2}) && (28)\\
&= P(\frac{\mu_{2i}}{\sigma_2}>-\frac{x_{2i}\beta_2}{\sigma_2}) && (29)\\
&= P(\frac{\mu_{2i}}{\sigma_2}<\frac{x_{2i}\beta_2}{\sigma_2}) && (30)\\
&= \Phi(x_{2i}\delta) && (31)
\end{aligned}$$

x_{2i} y y_{2i} se observan para todas las personas, δ se puede estimar por MV. No se puede identificar por separado β_2 y σ_2

Modelo de Heckman: Solución 4/4

Método en dos etapas

- Etapa 1: Estimar $\hat{\delta}$ usando un probit.

$$P(y_{2i} = 1) = \Phi(x_{2i}\delta)$$

  Se usa toda la muestra para obtener:

$$z_i = \lambda(\frac{-x_{2i}\beta_2}{\sigma_2})$$

- Etapa 2: Realizar la regresión de y_{1i} en x_{1i} y $\hat{z}_i$ usando la muestra bajo selección ($y_{2i} = 1$)

Aspectos finales

- El modelo El modelo de Heckman es un tipo de truncamiento, de hecho se le conoce también como truncamiento incidental.
- Debido al uso del Probit, el modelo se le conoce también como Heckit.
- La estimación de la varianza en la segunda etapa requiere una corrección, Stata ya lo considera.
- El modelo también puede ser estimado de manera conjunta por MV, aunque esto requiere asumir normalidad bivariada.
- Se puede probar la selección con $H0 : \delta = 0$ aunque, cuando x_1 es similar a x_2, la multicolinealidad alta afecta la potencia del test.

Stata

- Cálculo directo
  - generate d = (y < .)
  - heckman y x1, select(d x2)
  - heckman y x1, select(d x2) twostep
- A mano
  - generate d = (y < .)
  - probit d x2
  - predict xb, xb
  - generate lambda=normalden(xb)/normal(xb)
  - reg y x1 lambda if d==1

Contenido

Principales características:

- El modelo Poisson está asociado a valores no negativos enteros (espacio $\mathbb{Z}$) de la variable dependiente, además de tener una característica discreta y asociada a un conteo de casos.
- Por ejemplo, número de niños nacidos en el año; número de crímenes por semana; número de patentes por semana, número de casos Covid-19 en un día.

El modelo Poisson puede expresarse en el valor esperado de la función exponencial:

$$E(y|x_1, x_2, \dots, x_k) = \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k) \tag{32}$$

Taking logs

$$\log(E(y|x_1, x_2, \dots, x_k)) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \equiv \boldsymbol{x\beta} \tag{33}$$

La función Log-likelihood

La función de distribución Poisson está definida por la probabilidad condicional en λ para cada h:

$$P(Y = h|\lambda) = \frac{\exp(-\lambda)\lambda^h}{h!}$$

Así modelamos en cambio la función de likelihood $P(y = h|\lambda)$

$$P(Y = h|\lambda) = \frac{\exp(-\exp(\boldsymbol{x}\beta))\exp(\boldsymbol{x}\beta)^h}{h!}$$

taking logs:

$$\log P(Y = h|\lambda) = -\exp(\boldsymbol{x}\beta) + h\boldsymbol{x}\beta - \log(h!)$$

USAT Universidad Católica Santo Toribio de Mogrovejo

La función Log-likelihood

Considerando $h = y_i$ para una observación en particlar i.

$$\log P(Y = y_i|\lambda) = -\exp(\boldsymbol{x}\beta) + y_i\boldsymbol{x}\beta - \log(y_i!)$$

Sumando a través de las observaciones en la muestra:

$$L(\beta) = \sum_{i}^{n}\{y_i\boldsymbol{x}\beta - \exp(\boldsymbol{x}\beta) - \log(y_i!)\}$$

Los parámetros contenidos en el vector β se obtienen maximizando la función de log likelihood.

USAT Universidad Católica Santo Toribio de Mogrovejo

Partial effects

Calculando el porcentaje de cambio en el esperado condicional de y dado un cambio en la variable x (una semi-elasticidad):

$$\%\Delta(E(y|x_1, x_2, \ldots, x_k)) \approx 100\beta\Delta x_j \tag{34}$$

Siendo más precisos el cambio porcentual de $E(y|x)$ ante un cambio discreto en x

$$\left(\frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k(c_k + 1))}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k(c_k))} - 1\right) \times 100 \equiv \left(\exp(\beta_k) - 1\right) \times 100 \tag{35}$$

Donde el cambio discreto es igual a uno (1).

Partial effects

Se puede comparar los efectos parciales del modelo Poisson con el modelo logit y probit. Para esto se calcula $\frac{\partial E(y|x)}{\partial x_j}$ en la expresión (32):

$$\frac{\partial E(y|x)}{\partial x_j} = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)\beta_j \tag{36}$$

Tenemos que en el APE, el factor de escala es: $\sum_{i=1}^{n} \exp(\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik})$. Así entonces:

$$\frac{\partial E(y|x)}{\partial x_j} = \overline{y}\beta_j \tag{37}$$

Important issue

Aunque el modelo Poisson es una elección natural para modelar datos de conteo. Tiene un "important issue".

$$Var(y|x) = E(y|x) \tag{38}$$

En otras palabras es muy restrictivo. Aquí una pregunta de rigor: ¿Cuáles son las consecuencias? ¡No muy serias! Si la distibución Poisson "does o does not fit the data", los β's son aún consistentes y asintóticamente normal. Lo último es análogo al estimador OLS, el cual es consistente y asintóticamente normal si la suposición de normalidad se mantenga o no.

Important issue

Se puede hacer el modelo aún más flexible:

$$Var(y|x) = \sigma^2 E(y|x) \tag{39}$$

El parámetro σ^2 es desconocido. Si $\sigma^2 > 1$ entonces existe evidencia que el proceso generador de datos tiene una varianza de "overdispersion"; si $\sigma^2 < 1$ entonces la varianza tiene la característica de "underdispersion".

!Vayamos a STATA!

Exploremos la siguiente base de datos en la página de Wooldridge. Utilizar el siguientes comando "bcuse crime1". Una descripción de las variables en la siguiente figura:

```
CRIME1.DES

narr86    nfarr86   nparr86   pcnv      avgsen    tottime   ptime86   qemp86
inc86     durat     black     hispan    born60    pcnvsq    pt86sq    inc86sq

  Obs:   2725

  1. narr86                   # times arrested, 1986
  2. nfarr86                  # felony arrests, 1986
  3. nparr86                  # property crme arr., 1986
  4. pcnv                     proportion of prior convictions
  5. avgsen                   avg sentence length, mos.
  6. tottime                  time in prison since 18 (mos.)
  7. ptime86                  mos. in prison during 1986
  8. qemp86                   # quarters employed, 1986
  9. inc86                    legal income, 1986, $100s
 10. durat                    recent unemp duration
 11. black                    =1 if black
 12. hispan                   =1 if Hispanic
 13. born60                   =1 if born in 1960
 14. pcnvsq                   pcnv^2
 15. pt86sq                   ptime86^2
 16. inc86sq                  inc86^2
```

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

A modo de repaso

- Usamos probit y logit para una respuesta binaria.
- Usamos Tobit para una solución de esquina ("corner solution outcome")

Usualmente otro rasgo de los datos cae en la misma categoría de variables restringidas por un valor; en este caso estamos hablando de una variable censurada.

El modelo de datos censurados

¿Cómo surge una variable censurada?

- *Survey design*. Es un caso de missing data (en la variable dependiente). Por ejemplo, demanda por tickets en eliminatorias para los últimos cupos al mundial.
- En algunos casos restricciones institucionales (Wooldridge).

El modelo de datos censurados

- ¿Cúal es el rasgo del modelo de variable censurada? Las unidades son observables y proveen información de las variables independientes (X's); pero la información sobre la variable dependiente esta ausente (variable omitida). Es de conocimiento el valor de corte. Este valor puede ser superior (*right censoring*) o inferior (*left censoring*).
- Unidades escogen opciones como "mas de 50 000 dólares" (threshold); se observan datos menos de 50 000 dólares.
- La misma respuesta del *threshold* para muchas observaciones i.

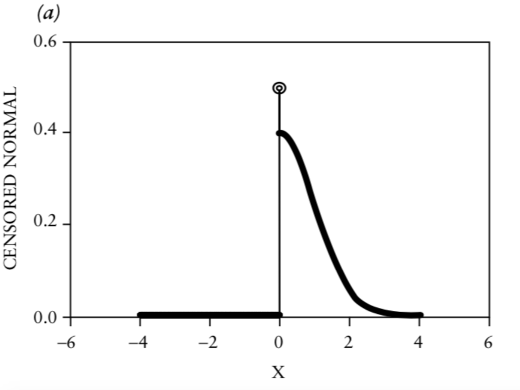


FIGURA: Source: Heij et. al 2004 “Econometric Methods with Applications in Business and Economics”

El modelo de datos censurados

El modelo censurado puede expresarse de la siguiente manera:

$$y_i = \boldsymbol{x_i}\beta + u_i, \;\; u_i|\; x_i, c_i \;\sim Normal(0, \sigma^2) \tag{40}$$
$$w_i = \min(y_i, c_i) \tag{41}$$

u_i is independent of c_i

¿Cuál es el problema de estimar este modelo por OLS? Las razones son similares como en el caso del modelo TOBIT.

- Los β's son inconsistentes.

Sin embargo existe un punto importante; en el modelo TOBIT, se esta modelando comportamiento óptimo de los individuos ($y \equiv$ consumo de alcohol), y en el modelo censurado se tiene un problema en el método de recolección porque (por alguna razón) una porción de los datos son censurados (no observados).

El modelo de datos censurados

Exploremos la siguiente base de datos en la página de Wooldridge. Utilizar el siguientes comando “bcuse recid”. Una descripción de las variables en la siguiente figura:

```
RECID.DES

black     alcohol   drugs     super     married   felon     workprg   property
person    priors    educ      rules     age       tserved   follow    durat
cens      ldurat

  Obs:  1445

  1. black                   =1 if black
  2. alcohol                 =1 if alcohol problems
  3. drugs                   =1 if drug history
  4. super                   =1 if release supervised
  5. married                 =1 if married when incarc.
  6. felon                   =1 if felony sentence
  7. workprg                 =1 if in N.C. pris. work prg.
  8. property                =1 if property crime
  9. person                  =1 if crime against person
 10. priors                  # prior convictions
 11. educ                    years of schooling
 12. rules                   # rules violations in prison
 13. age                     in months
 14. tserved                 time served, rounded to months
 15. follow                  length follow period, months
 16. durat                   max(time until return, follow)
 17. cens                    =1 if duration right censored
 18. ldurat                  log(durat)
```

Figura: recid.des in Wooldridge’s datasets

El modelo de datos truncados

- Un modelo de datos censurados es aplicable cuando se tiene observaciones de las unidades y la información es parcial; es decir, se tiene información de las variables independientes pero no de la variable y.
- Usamos Tobit para una solución de esquina ("corner solution outcome").
- Un modelo truncado tiene la característica de excluir (basado en el valor de y). No se tiene una submuestra aleatoria. Sin embargo, tenemos conocimiento de la regla de exclusión. Esta regla esta determinada si y esta por encima o por debajo de cierto valor (o "threshold").

El modelo de datos truncados

¿Como surge o se identifica un modelo de datos truncados?

- El investigador presta atención a una submuestra de la población (quizas debido a costos de muestreo, ver Wooldridge).
- Hay que enfatizar que la estimación por OLS es eficiente cuando la muestra seleccionada es aleatoria.

Ejemplos: Hausman y White (1977) usan data de impuestos negativos a la renta como determinante de las ganancias individuales/familiares. El estudio solo incluía familias con renta 1.5 veces la linea de pobreza.

El modelo de datos truncados

El modelo de datos truncados puede expresarse de la siguiente manera;

$$y = \boldsymbol{x}\beta + u, u \mid x, \sim Normal(0, \sigma^2) \tag{42}$$

y el set de datos (y, x) es observado solo si $y \geq c_i$ donde el "threshold" depende de variables x. Por ejemplo, Haussman y white (1977) definen c_i como el tamaño de la familia.

La funcion de distribución gráficamente luce de la siguiente manera;

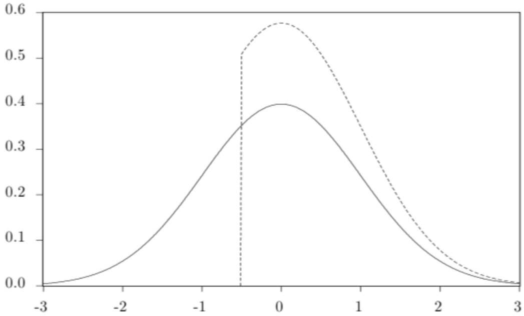


FIGURA: Source: Chumacero, R, 2003: “Limited dependent variable” handout

El modelo de datos truncados

La funcion de distribución formalmente luce de la siguiente manera;

$$g(y|x_i, c_i) = f(y|x_i\beta, \sigma^2)/F(c_i|x_i, \beta, \sigma^2), y \leq c_i \tag{43}$$

donde $f(y|x_i\beta, \sigma^2)$ denota la función de densidad normal y $F(c_i|x_i, \beta, \sigma^2)$ es la función acumulada evaluada en el threshold c_i. En pocas palabras, se re-pondera dividiendo la función de densidad normal por la acumulada para que la nueva función de densidad sume el valor de uno (1) sobre el dominio de los datos. Luego, se toman logs y se estima por ML (es decir, se maximiza la función ($g(y|x_i, c_i)$ con los datos observados).

USAT Universidad Católica Santo Toribio de Mogrovejo

¡Vayamos a STATA!

Tengamos en cuenta un estudio que tiene el objetivo de modelar el desempeño académico como una función de las destrezas de lenguaje ("language skills") y el tipo de programa en el que los estudiantes se han matriculado. Se requiere analizar a los estudiantes que tengan un desempeño mínimo de 40. Tengamos en cuenta la siguiente base de datos en formato STATA:
use https://stats.idre.ucla.edu/stat/stata/dae/truncreg, clear

Contenido

Inferencia

En todos los modelos mostrados anteriormente se describe explícitamente la función de log-likelihood. Bajo condiciones regulares el MLE es consistente y asintóticamente normal. La matriz de varianzas y covarianzas puede ser estimada calculando la inversa de la matriz de información:

$$I(\beta)^{-1} = -E\left[\frac{\partial^2 L(\beta)}{\partial\beta\partial\beta}\right]$$

Alternativamente,

$$\frac{\partial L(\beta)}{\partial\beta}\frac{\partial L(\beta)}{\partial\beta'}$$

Hay que recordar que en ambos casos se asume que el modelo está especificado correctamente. Si lo anterior no es el caso, un estimador asintótico de la matriz de varianza y covarianzas es requerido.

- El modelo de datos truncados es un buen ejemplo de la selección no aleatoria de la muestra.
- Sin embargo, el diseño de investigación puede incluir una muestra (representativa) de la población, y aún tienes acceso o disponibilidad de una sub-muestra –los entrevistados rehusan responder al "survey" por ejemplo.
- Cuando lo anterior ocurre, la pregunta inmediata es: ¿cómo esto afecta a mis estimadores en términos de eficiencia y sesgo?

Vamos el caso de "incidental truncation". Antes revisaremos la implicancia en cuestión de sesgo y consistencia.

Contenido

- consideremos el siguiente modelo:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u, \quad E(u|x_1, x_2, \ldots, x_k) = 0 \quad (44)$$

re-escribiendo el modelo anterior;

$$y = \boldsymbol{x_i}\boldsymbol{\beta} + u \quad (45)$$

Tengamos el siguiente indicador S que toma el valor 1 si se observa la dupla y_i, x_i y cero de otra manera.

$$s_i y_i = s_i \boldsymbol{x_i}\boldsymbol{\beta_i} + s_i u \quad (46)$$

Consistencia se consigue si el termino del error tiene esperanza cero y es no está correlacionado con las x's.

$$E[(sx_j)(su)] \equiv E[sx_j u] = 0 \tag{47}$$

Una condición clave para el insesgamiento (lo cual no es requerido para consistencia) es $E(su|x) = 0$. Si la selección de la muestra es completamente aleatoria entonces i.e. x y s son independientes $-E(sx_j u) \equiv E(s)E(x_j u) = 0$. En otras palabras, si los datos se excluye de manera aleatoria, OLS es consistente e insesgado.

- En general, si s depende de variables explicativas y términos aleatorios adicionales que son independientes de x y u, OLS es consistente e insesgado.
- Por ejemplo, IQ es una variable explicativa en un modelo de horas de trabajo / salario. Pero la información de IQ es limitada; $s = 1$ si IQ$\geq \nu$ y s=0 si IQ$< \nu$. El parámetro ν no es fijo y es una variable no observada que es independiente de IQ, u y de otras variables explicativas. Asi s es una variable aleatoria (sin patrón definido) como por ejemplo una bernoulli.
- En términos de eficiencia, si s es independiente de u, entonces $E(u|x,s) = E(u|x)$. Así entonces, si se añade la hipótesis de homocedasticidad $E(u^2|x,s) = E(u^2) = \sigma^2$. Luego los errores estándar son válidos.

Contenido

¿Cuando OLS en la muestra es inconsistente?

- En un modelo de truncamiento, se observa que $s = 1$ si $y_i \leq c_i$ donde c_i es el threshold que define el truncamiento.
- Equivalentemente $s_i = 1$ si $u_i \leq c_i - x_i\beta$. Porque s_i depende directamente de u_i, entonces s_i y u_i están correlacionados. Esta es la razón por la cual OLS sobre una muestra seleccionada no entrega estimadores de β consistentes.

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

Incidental truncation

A veces la regla que hace visualizar el resultado y no es exógena. Por ejemplo, en el caso del salario (por hora), las personas fuera de la fuerza laboral no hacen visible un salario. Así el truncamiento de la variable de salario es "incidental" por que esta depende de otra variable relacionada a la participación laboral. Otras variables pueden ser observadas como educación, experiencia, género, estado marital y demás.
El modelo que puede lidiar con el "incidental truncation" tiene la siguiente forma:

$$y = \boldsymbol{x}\beta + u, \ \ E(u|x, z) = 0 \tag{48}$$

$$s = 1[\boldsymbol{z}\gamma + \nu \geq 0] \tag{49}$$

Incidental truncation

El modelo a estimar es:

$$E(y|z, s = 1) = \boldsymbol{x}\beta + \rho\lambda(\boldsymbol{z}\boldsymbol{\gamma}) \quad (50)$$
$$P(s = 1|\boldsymbol{z}) = \Phi(\boldsymbol{z}\gamma) \quad (51)$$

Es decir, se estima primero $\boldsymbol{\gamma}$ con un probit ($\boldsymbol{s}$ sobre $\boldsymbol{z}$ utilizando la muestra completa. Luego, podemos estimar $\boldsymbol{\beta}$. Pregunta de examen: ¿cómo se llama el procedimiento de identificación: Los dos pasos de Heckman (o Heckit).

Incidental truncation

Utilicemos la base de datos “mroz.dta” en la estimación de un Heckit.

Contenido

USAT Universidad Católica Santo Toribio de Mogrovejo

REFERENCIAS

- Referencia 1
- Referencia 2. *colocar alguna referencia*
- Referencia 3. *colocar alguna referencia*

2021-09-24

Capítulo 17
└─Referencias

└─Referencias

Referencia 1
Referencia 2. *colocar alguna referencia*
Referencia 3. *colocar alguna referencia*

- Agregar alguna nota

Econometría Básica

Capítulo 17: Variable Limitada

José Valderrama & Freddy Rojas
jtvalderrama@gmail.com & frojasca@gmail.com
Universidad Católica Santo Toribio de Mogrovejo

Septiembre de 2021