Estimación Tobit en gretl

Microeconomía Cuantitativa

R. Mora

Departmento de Economía
Universidad Carlos III de Madrid

Esquema

El modelo Tobit y la estimación ML

El modelo Tobit

- $h^* = \beta x + \varepsilon$
- $\varepsilon \sim N\left(0, \sigma^2\right)$
- $\begin{cases} \text{if } h^* > 0 \Rightarrow h = \beta x + \varepsilon \\ \text{if } h^* \leq 0 \Rightarrow h = 0 \end{cases}$

$$\hat{\beta}^{ML} = \arg\max \quad \sum_i \Big\{ 1\left(h_i > 0\right) \log\left(\left(\tfrac{1}{\sigma}\right) \phi\left(\tfrac{h_i - \beta x_i}{\sigma}\right)\right) \\ + 1\left(h_i = 0\right) \log\left(1 - \Phi\left(\tfrac{\beta x_i}{\sigma}\right)\right)\Big\}$$

- en gretl, se emplea un algoritmo quasi-Newton (el algoritmo BFGS)

Comandos básicos en gretl para la estimación Tobit

- `tobit`: computa la estimación del modelo Tobit por ML
- `omit/add`: contrasta la significatividad conjunta
- `$yhat`: estima la variable dependiente
- `$lnl`: computa la log-verosimilitud del último modelo estimado
- `intreg`: computa la estimación ML del modelo de regressión con variable censurada en intervalos (interval regression model) y errores normales

- en esta Sesión vamos a aprender a utilizar `tobit` y `intreg`

tobit *depvar indvars* --verbose

- "censura": la variable dependiente toma valores no-negativos, con el 0 reservado para las observaciones censuradas
- para otros casos (censura por arriba, o un punto de masa differente a 0) se debe redefinir la variable dependiente
- para censura por dos lados, se puede utilizar `intreg`
- es posible ver todas las iteraciones con `--verbose`
- el output muestra el test χ^2_q de la nula que todas las pendientes son iguales a 0
- por defecto, las desviaciones típicas se calculan con la inversa negativa del Hessiano

Ejemplo: datos simulados

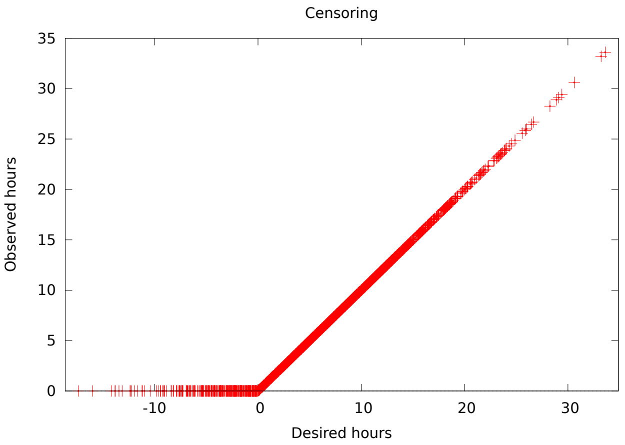
El modelo Tobit

- $h^* = 10 + 0.5 * educ - 5 * kids + \varepsilon$
- $\varepsilon \sim N(0, 49)$

- la educación te hace querer trabajar más horas
- tener niños te hace querer trabajar menos horas
- $\beta x = 10 + 0.5 * educ - 5 * kids$
- $\sigma = 7$

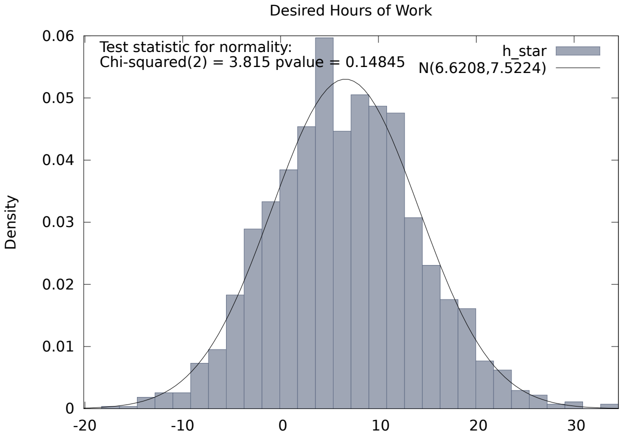
Horas Observadas vs. Horas Deseadas

$$h^* = 10 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$



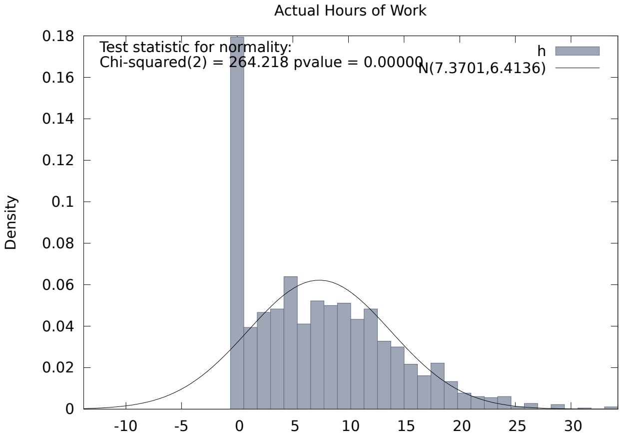
Distribución de Horas Deseadas

$$h^* = 10 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$



Censura en el modelo Tobit

$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$



ols con muestra completa

$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0,49)$$

```
Model 1: OLS, using observations 1-1500
Dependent variable: h

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const        6.60589     0.757742       8.718   7.39e-18 ***
  educ         0.387944    0.0751737      5.161   2.79e-07 ***
  kids        -4.48766     0.382296     -11.74    1.68e-30 ***

Mean dependent var   7.370101   S.D. dependent var   6.413578
Sum squared resid    54501.48   S.E. of regression   6.033833
R-squared            0.116094   Adjusted R-squared   0.114913
F(2, 1497)           98.30967   P-value(F)           7.67e-41
Log-likelihood      -4822.980   Akaike criterion     9651.960
Schwarz criterion    9667.899   Hannan-Quinn         9657.898
```

ols con muestra restringida

$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$

```
Model 2: OLS, using observations 1-1223
Dependent variable: h

             coefficient   std. error   t-ratio   p-value
  --------------------------------------------------------
  const        7.35385     0.773674      9.505    1.02e-20 ***
  educ         0.333273    0.0764336     4.360    1.41e-05 ***
  kids        -2.84425     0.389046     -7.311    4.79e-13 ***

Mean dependent var   9.134221   S.D. dependent var   5.678892
Sum squared resid    37539.67   S.E. of regression   5.547091
R-squared            0.047441   Adjusted R-squared   0.045879
F(2, 1220)           30.38012   P-value(F)           1.33e-13
Log-likelihood      -3829.194   Akaike criterion     7664.387
Schwarz criterion    7679.715   Hannan-Quinn         7670.156
```

La salida de tobit

$$h^* = 5 + 0.5 * educ - 5 * kids + \varepsilon, \varepsilon \sim N(0, 49)$$

```
Function evaluations: 48
Evaluations of gradient: 9

Model 3: Tobit, using observations 1-1500
Dependent variable: h

             coefficient   std. error     z       p-value
  -------------------------------------------------------
  const        4.89145     0.893211      5.476   4.34e-08 ***
  educ         0.461985    0.0891748     5.181   2.21e-07 ***
  kids        -4.44067     0.470267     -9.443   3.63e-21 ***

Mean dependent var   7.447435   S.D. dependent var   6.233858
Censored obs              277   sigma                7.089319
Log-likelihood      -4412.070   Akaike criterion     8832.141
Schwarz criterion    8853.394   Hannan-Quinn         8840.058

Test for normality of residual -
  Null hypothesis: error is normally distributed
  Test statistic: Chi-square(2) = 3.57203
  with p-value = 0.167627
```

Prediciendo las horas de trabajo entre los que trabajan

computando $\hat{h}_i^*$ y $\hat{h}_i$

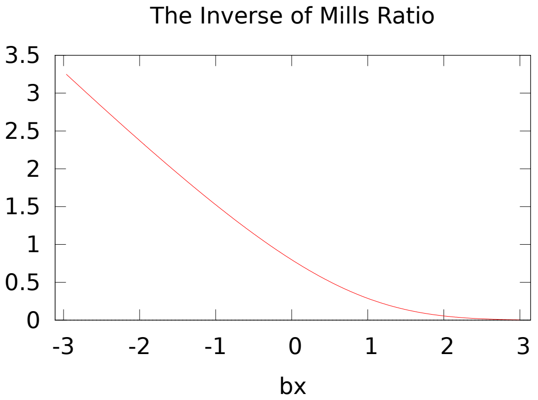
- $\hat{h}_i^*$: genr h_star_hat=$yhat
- para cada i, $\hat{h}_i = \max\left\{0, \hat{\beta}x_i\right\}$

$E\left[h \,|\, h > 0, x\right]$

- $E\left[h \,|\, h > 0, x\right] = \beta x + E\left[\varepsilon \,|\, \beta x + \varepsilon > 0, x\right]$
- se pueded demostrar que: $E\left[h \,|\, h > 0, x\right] = \beta x + \sigma \frac{\phi(\frac{\beta x}{\sigma})}{\Phi(\frac{\beta x}{\sigma})}$

- $\frac{\phi(\frac{\beta x}{\sigma})}{\Phi(\frac{\beta x}{\sigma})}$ es la inversa del ratio de Mills (John Mills en realidad usó $\frac{1-\Phi(\frac{\beta x}{\sigma})}{\phi(\frac{\beta x}{\sigma})}$)

La inversa del ratio de Mills



a mayor $\frac{\beta x}{\sigma}$, menor necesidad de corrección (menor probabilidad de censura)

Prediciendo las horas observadas de trabajo

$E[h|x]$

- $E[h|x] = \Pr(h>0)\,E[h|h>0,x]$
- se puede demostrar que: $E[h|x] = \Phi\left(\frac{\beta x}{\sigma}\right)\left[\beta x + \sigma\frac{\phi(\beta x)}{\Phi(\beta x)}\right]$

Entendiendo las pendientes

- las estimaciones de las pendientes, $\hat{\beta}$, nos dan los efectos marginales del número deseado de horas
- a veces, queremos estimar el efecto sobre la probabilidad de trabajar y sobre el número real de horas trabajadas

Efectos marginales algebráicos

Probabilidad de participar

- $\frac{\partial \Pr(h_i > 0)}{\partial x_j} = \phi(\frac{\beta x}{\sigma}) \left(\frac{\beta_j}{\sigma}\right)$

Horas realmente trabajadas

- $\frac{\partial E(h_i|x)}{\partial x_j} = \beta_j \Phi\left(\frac{\beta x}{\sigma}\right)$
- se pueden obtener aproximaciones con OLS con la muestra completa

Efectos marginales individuales: un cambio discreto

- a vece queremos computar los efectos individuales

Cambio discreto

- almacene los coeficientes estimados en un vector
- genere la matriz con los controles bajo el escenario 0, x_0, y la matriz del escenario 1, x_1
- prediga $\hat{\beta}^{ML}x_0$ y $\hat{\beta}^{ML}x_1$
- simule la censura
- genere los efectos marginal individuales

Ejemplo: el efecto de tener un niño adicional

```
# numerical individual marginal effects: having an extra kid
genr beta=$coeff
genr kids1=kids+1
matrix x0={const,educ,kids}
matrix x1={const,educ,kids1}
series h0 = (x0*beta>0)*x0*beta
series h1 = (x1*beta>0)*x1*beta
series Mg_kid=h1-h0
summary Mg_kid --by=educ --simple
```

summary Mg_kid --by=educ --simple

también el efecto es menor con mayor educación

	Mean	Minimum	Maximum	Std. Dev.
educ = 8	-9.4738	-9.4738	-9.4738	0.0000
educ = 12	-7.1477	-11.164	-5.9182	2.2237
educ = 16	-5.0951	-7.6081	-2.3627	2.6231
educ = 21	-4.4750	-4.4750	-4.4750	0.0000

El modelo Interval Regression

$m_i \leq x_i\beta + \varepsilon_i \leq M_i$

- de la variable dependiente solo se observa que pertenece a un intervalo

- en la práctica, cada observación pertenece a uno de los siguientes casos:
  - acotada por la izquierda
  - acotada por la derecha
  - acotada
  - observación puntual

El modelo Interval Regression en gretl

intreg minvar maxvar X

- minvar contiene m_i, con NAs para observaciones no-acotadas por la izquierda
- maxvar contiene M_i, con NAs para observaciones no-acotadas por la derecha
- puede que m_i sea igual a M_i para algunas observaciones
- los errores estándar se computan usando la inversa negativa del Hessiano
- si la flag −−robust está activada, estimadores sandwich de los errores estándar
- es fácil computar el ratio de verosimilitudes

Un ejemplo

```
nulldata 100
# generate artificial data
set seed 201449
x = normal()
epsilon = 0.2*normal()
ystar = 1 + x + epsilon
lo_bound = floor(ystar)
hi_bound = ceil(ystar)
# run the interval model
intreg lo_bound hi_bound const x
```

La salida intreg

```
Model 1: Interval estimates, using observations 1-100
Lower limit: lo_bound, Upper limit: hi_bound

             coefficient   std. error     z      p-value
  --------------------------------------------------------
  const       0.977619     0.0360159    27.14   2.97e-162 ***
  x           0.993708     0.0375009    26.50   1.02e-154 ***

Chi-square(1)        702.1575   p-value              1.0e-154
Log-likelihood      -40.45572   Akaike criterion     86.91144
Schwarz criterion    94.72695   Hannan-Quinn         90.07452

sigma = 0.225354
Left-unbounded observations: 0
Right-unbounded observations: 0
Bounded observations: 100
Point observations: 0

Test for normality of residual -
  Null hypothesis: error is normally distributed
  Test statistic: Chi-square(2) = 27.391
  with p-value = 1.1275e-06
```

Resumen

- gretl permite estimación ML para el modelo Tobit
- el modelo Tobit identifica cómo afecta cada control la probabilidad de participar y, si participa, el efecto marginal sobre la variable dependiente
- el modelo interval regression por ML también se puede estimar en gretl