Microeconometría I

Maestría en Econometría

Lecture 5

Tobit and Selection Models

1. Tobit and Selection Models
   - Introduction
   - Censoring and Truncation Mechanisms
   - Censored and Truncated MLE
   - Tobit Model
2. Sample Selection Models
3. La Oferta de Trabajo de las Mujeres

Agenda

1. Tobit and Selection Models
   - Introduction
   - Censoring and Truncation Mechanisms
   - Censored and Truncated MLE
   - Tobit Model

2. Sample Selection Models

3. La Oferta de Trabajo de las Mujeres

Introduction

- We consider two closely related topics: regression when the dependent variable of interest is incompletely observed and regression when the dependent variable is completely observed but is observed in a selected sample that is not representative of the population.
- This includes limited dependent variable models, latent variable models, generalized Tobit models, and selection models.

Introduction

- All these models share the common feature that even in the simplest case of population conditional mean linear in regressors, OLS regression leads to inconsistent parameter estimates because the sample is not representative of the population.
- Leading causes of incompletely observed data are truncation and censoring
- For truncated data some observations on both the dependent variable and regressors are lost. For example, income may be the dependent variable and only low-income people are included in the sample.

Introduction

- For censored data information on the dependent variable is lost, but not data on the regressors. For example, people of all income levels may be included in the sample, but for confidentiality reasons the income of high-income people may be top-coded and reported only as exceeding, say, \$100,000 per year.
- A leading example of truncation and censoring is the Tobit model, named after Tobin (1958), who considered linear regression under normality.

Introduction

- Let y^* denote a variable that is incompletely observed.
- For truncation from below, y^* is only observed if y^* exceeds a threshold.
- For simplicity, let that threshold be zero.
- Then we observe $y = y^*$ if $y^* > 0$.
- Since negative values do not appear in the sample, the truncated mean exceeds the mean of y^*.

Introduction

- Let y^* denote a variable that is incompletely observed.
- For censoring from below at zero, y^* is not completely observed when $y^* = 0$, but it is known that $y^* < 0$ and for simplicity y is then set to 0.
- Since negative values are scaled up to zero, the censored mean also exceeds the mean of y^*.
- Clearly, sample means in truncated or censored samples cannot be used without adjustment to estimate the original population mean.

Introduction

- With luck, truncation and censoring might lead only to a shift up or down in the intercept, leaving slope coefficients unchanged; however, this is not the case.
- For example, if $E[y^*|x] = x\beta$ in the original model then truncation or censoring leads to $E[y|x]$ being nonlinear in x and β so that OLS gives inconsistent estimates of β and hence inconsistent estimates of marginal effects.

Example

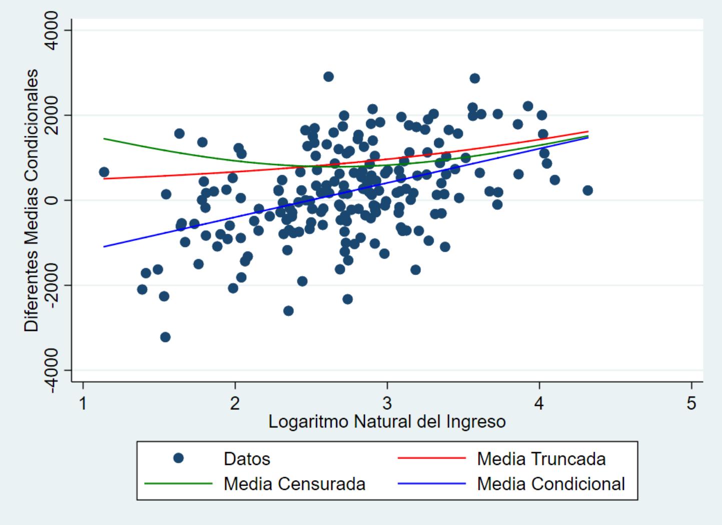
- As an illustration we consider the following labor supply example with simulated data.
- The relationship between desired annual hours worked, y^*, and hourly wage, w, is specified to be of linear-log form with data-generation process:

$$y^* = -2500 + 1000 \ln w + \epsilon,$$

$$\epsilon \sim \text{Normal}(0, 1000^2)$$

$$\ln w \sim \text{Normal}(2.75, 0.60^2)$$

Example


Diferentes Medias Condicionales
4000
2000
0
-2000
-4000
1
2
3
4
5
Logaritmo Natural del Ingreso
● Datos
Media Truncada
Media Censurada
Media Condicional

Example

- It is clear that censored and truncated conditional means are nonlinear in x even if the underlying population mean is linear.
- OLS estimation using truncated or censored data will lead to inconsistent estimation of the slope parameter, since by visual inspection of last figure a linear approximation to the nonlinear truncated and censored means will have flatter slope than that for the original untruncated mean.

Agenda

1. Tobit and Selection Models
   - Introduction
   - Censoring and Truncation Mechanisms
   - Censored and Truncated MLE
   - Tobit Model
2. Sample Selection Models
3. La Oferta de Trabajo de las Mujeres

Censoring and Truncation Mechanisms

- Let y denote the observed value of the dependent variable.
- The departure from usual analysis is that y is the incompletely observed value of a latent dependent variable y^*, where the observation rule is

$$y = g(y^*),$$

for some specified function $g(\cdot)$.

Censoring

- With censoring we always observe the regressors x, completely observe y^* for a subset of the possible values of y^*, and incompletely observe y for the remaining possible values of y^*.
- If censoring is from below (or from the left), we observe

$$y = \begin{cases} y^* & \text{If } y^* > L \\ L & \text{If } y^* \leq L \end{cases}$$

- Example: all consumers may be sampled with some having positive durable goods expenditures ($y^* > 0$) and others having zero expenditures ($y^* = 0$).

Censoring

- If censoring is from above (or from the right) we observe

$$y = \begin{cases} y^* \text{ If } y^* < U \\ U \ \text{ If } y^* \geq U \end{cases}$$

- Example: annual income data may be top-coded at U $=$ \$100,000.

Truncation

- Truncation entails additional information loss as all data on observations at the bound are lost.
- With truncation from below we observe only

$$y = y^* \text{ if } y^* > L.$$

- Example: only consumers who purchased durable goods may be sampled $(L = 0)$.

Truncation

- With truncation from above we observe only

$$y = y^* \text{ if } y^* < U.$$

- Example, only low-income individuals may be sampled.

Agenda

1. Tobit and Selection Models
   - Introduction
   - Censoring and Truncation Mechanisms
   - Censored and Truncated MLE
   - Tobit Model
2. Sample Selection Models
3. La Oferta de Trabajo de las Mujeres

Censored and Truncated MLE

- If the conditional distribution of y^* given regressors x is specified, then the parameters of this distribution can be consistently and efficiently estimated by ML estimation based on the conditional distribution of the censored or truncated y.
- Let $f^*(y^*|x)$ and $F^*(y^*|x)$ denote the conditional probability density function (or probability mass function) and cumulative distribution function of the latent variable y^*.

Censored and Truncated MLE

- One can always obtain $f(y|x)$ and $F(y|x)$, the corresponding conditional pdf and cdf of the observed dependent variable y, since $y = g(y^*)$ is a transformation of y^*.
- Consider ML estimation given censoring from below.
- For $y > L$ the density of y is the same as that for y^*, so $f(y|x) = f^*(y|x)$.
- For $y = L$, the lower bound, the density is discrete with mass equal to the probability of observing $y^* \leq L$, or $F^*(L|x)$.

Censored MLE

- Thus for censoring from below

$$f(y|x) = \begin{cases} f^*(y|x) & \text{If } y > L \\ F^*(L|x) & \text{If } y = L \end{cases}$$

- Similar to analysis for binary outcome models, it is notationally convenient to introduce an indicator variable

$$d = \begin{cases} 1 & \text{If } y > L \\ 0 & \text{If } y = L \end{cases} \tag{1}$$

Censored MLE

- Then the conditional density given censoring from below can be written as

$$f(y|x) = f^*(y|x)^d F^*(L|x)^{1-d}$$

- For a sample of N independent observations, the censored MLE maximizes

$$\ln L_N(\theta) = \sum_{i=1}^{N} \{d_i \ln f^*(y_i|x_i, \theta) + (1 - d_i) \ln F^*(L_i|x_i, \theta)\} \tag{2}$$

Censored MLE

- For generality the censoring lower bound L_i is permitted to vary across individuals, though usually $L_i = L$.
- The censored MLE is consistent and asymptotically normal, provided the original density of the uncensored variable $f^*(y^*|x, \theta)$ is correctly specified.
- When censoring is instead from above, the log-likelihood is similar to (2), except now $d = 1$ if $y < U$ and $d = 0$ otherwise, and $F^*(L|x, \theta)$ is replaced by $1 - F^*(U|x, \theta)$.

Truncated MLE

- For truncation from below at L, and suppressing dependence on x, the conditional density of the observed y is

$$\begin{aligned} f(y) &= f^*(y|y > L) \\ &= f^*(y)/Pr[y|y > L] \\ &= f^*(y)/[1 - F^*(L)]. \end{aligned}$$

- The truncated MLE therefore maximizes

$$\ln L_N(\theta) = \sum_{i=1}^{N} \{\ln f^*(y_i|x_i, \theta) - \ln [1 - F^*(L_i|x_i, \theta)]\} \tag{3}$$

Truncated MLE

- If instead truncation is from above, the log-likelihood is (3), except that $[1 - F^*(L|x, \theta)]$ is replaced by $F^*(U|x, \theta)$.
- Ignoring censoring or truncation leads to inconsistency.
- If truncation is ignored the MLE maximizes

$$\sum_i \ln f^*(y_i|x_i, \theta),$$

which is the wrong likelihood function as it drops the second term in (3).

Poisson Truncated MLE Example

- Assume that y^* is Poisson distributed, so that $f^*(y) = e^{-\mu}\mu^y/y!$ and $\ln f^*(y) = -\mu + y \ln \mu - \ln y!$, with mean $\mu = \exp(x'\beta)$.
- Suppose the number of cigarettes smoked is modeled, but data are only available for people who smokes.
- Then the data are truncated from below at zero and we only observe $y = y^*$ if $y^* > 0$.
- Then $F^*(0) = Pr[y^* \leq 0] = Pr[y^* = 0] = e^{-\mu}$.

Poisson Truncated MLE Example

- From (3) the truncated MLE for β maximizes

$$\begin{aligned} \ln L_N(\beta) &= \sum_{i=1}^{N} \{- \exp x_i'\beta + y x_i'\beta - \ln y! \\ &- \ln\left[1 - \exp\left(- \exp\left(x_i'\beta\right)\right)\right]\} \end{aligned}$$

Poisson Truncated MLE Example

- Suppose instead that data are censored from above at 10 because of top-coding, so that we observe $y = y^*$if $y^* < 10$ and that $y = 10$ if $y^* \geq 10$. Then $Pr[y^* \geq 10] = 1 - Pr[y^* < 10] = 1 - \sum_{k=0}^{9} f^*(k)$. Then the censored MLE for β maximizes

$$
\begin{aligned}
\ln L_N(\beta) &= \sum_{i=1}^{N} \{ d_i [- \exp x_i'\beta + y x_i'\beta - \ln y!] \\
&+ (1 - d_i) \ln [\sum_{k=0}^{9} \exp(- \exp(x_i'\beta))(\exp x_i'\beta)^k / k!] \Big\}
\end{aligned}
$$

Agenda

1. Tobit and Selection Models
   - Introduction
   - Censoring and Truncation Mechanisms
   - Censored and Truncated MLE
   - Tobit Model
2. Sample Selection Models
3. La Oferta de Trabajo de las Mujeres

Tobit Model

- Truncation and censoring arise most often in econometrics in the linear regression model with normally distributed error, when only positive outcomes are completely observed.
- This model is called the Tobit model after Tobin (1958), who applied it to individual expenditures on consumer durable goods.

Tobit Model

- The censored normal regression model, or Tobit model, is one with censoring from below at zero where the latent variable is linear in regressors with additive error that is normally distributed and homoskedastic.

$$y^* = x'\beta + \epsilon, \tag{4}$$

  where

$$\epsilon \sim \text{Normal}[0, \sigma^2]$$

- This implies that the latent variable $y^* \sim N[x'\beta, \sigma^2]$.

Tobit Model

- The observed y is defined by

$$y = \begin{cases} y^* \text{ If } y^* > L \\ L \ \text{ If } y^* \leq L \end{cases}$$

  with $L = 0$.

- The conditional density given above is

$$f(y|x) = f^*(y|x)^d F^*(L|x)^{1-d}$$

Tobit Model

- with $f^*(y|x) \sim N[x'\beta, \sigma^2]$ and

$$\begin{aligned} F^*(0|x) &= Pr[y^* \leq 0] \\ &= Pr[x'\beta + \epsilon \leq 0] \\ &= \Phi(-x'\beta/\sigma) \\ &= 1 - \Phi(x'\beta/\sigma) \end{aligned}$$

- The censored density can be expressed as

$$f(y) = \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{1}{2\sigma^2}(y - x'\beta)^2\}\right]^d \left[1 - \Phi(\frac{x'\beta}{\sigma})\right]^{1-d} \quad (5)$$

Tobit Model

- The Tobit MLE $\hat{\theta} = (\hat{\beta}', \hat{\sigma}^2)'$ maximizes the censored log-likelihood function (2).
- Given (5) this becomes

$$\begin{aligned} \ln L_N(\beta, \sigma^2) &= \sum_{i=1}^{N} \left\{ d_i \left(-\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} (y_i - x_i'\beta)^2 \right) \right. \\ &- \left. (1 - d_i) \ln \left(\Phi(\frac{x'\beta}{\sigma}) \right) \right\}, \qquad (6) \end{aligned}$$

Tobit Model

- The first-order conditions are

$$\frac{\partial \ln L_N}{\partial \beta} = \sum_{i=1}^{N} \frac{1}{\sigma^2}\left(d_i(y_i - x_i'\beta) - (1-d_i)\frac{\sigma\phi_i}{1-\Phi_i}\right)x_i = 0 \tag{7}$$

$$\frac{\partial \ln L_N}{\partial \sigma^2} = \sum_{i=1}^{N}\left\{d_i\left(\frac{1}{2\sigma^2} + \frac{(y_i - x_i'\beta)^2}{2\sigma^4}\right) + (1-d_i)\frac{\phi_i x_i'\beta}{1-\Phi_i}\frac{1}{2\sigma^3}\right\} = 0 \tag{8}$$

where $\phi_i = \phi(x_i'\beta/\sigma)$ and $\Phi_i = \Phi(x_i'\beta/\sigma)$.

Tobit Model

- A very major weakness of the Tobit MLE is its heavy reliance on distributional assumptions.
- If the error ϵ is either heteroskedastic or nonnormal the MLE is inconsistent.
- Consistent estimation with heteroskedastic normal errors is possible by specifying a model for heteroskedasticity, say $\sigma_i^2 = exp(z_i'\gamma)$.
- Consistency then requires normal errors and correct specification of the functional form of the heteroskedasticity.

Tobit Model

- If data are truncated, rather than censored, from below at zero then the Tobit MLE $\hat{\theta} = (\hat{\beta}', \hat{\sigma}^2)'$ maximizes the truncated normal log-likelihood function

$$\begin{aligned} \ln L_N(\beta, \sigma^2) &= \sum_{i=1}^{N} \left\{ -\frac{1}{2} \ln 2\pi - \frac{1}{2} \ln \sigma^2 - \frac{1}{2\sigma^2}(y_i - x_i'\beta)^2 \right. \\ &- \left. \ln \left(1 - \Phi(\frac{x'\beta}{\sigma}) \right) \right\}, \qquad (9) \end{aligned}$$

Censored and Truncated Means in Linear Regression

- Censoring and truncation in the linear regression model (4) lead to observed dependent variable y that:
  1. has distribution with conditional mean other than $x'\beta$,
  2. conditional variance other than σ^2 even if ϵ is homoskedastic,
  3. and distribution that is nonnormal even if ϵ is normally distributed.
- Truncated mean. The effects of truncation are intuitively predictable. Left-truncation excludes small values, so the mean should increase, whereas with right-truncation the mean should decrease.
- Since truncation reduces the range of variation, the variance should decrease.

Censored and Truncated Means in Linear Regression

- For left-truncation at zero we only observe y if $y^* > 0$. If we suppress dependence of expectations on x for notational simplicity, the left-truncated mean becomes

$$\begin{aligned} E[y] &= E[y^*|y^* > 0] \\ &= E[x'\beta + \epsilon | x'\beta + \epsilon > 0] \\ &= E[x'\beta | x'\beta + \epsilon > 0] + E[\epsilon | x'\beta + \epsilon > 0] \\ &= x'\beta + E[\epsilon|\epsilon > -x'\beta] \end{aligned} \tag{10}$$

- As expected the truncated mean exceeds $x'\beta$, since $E[\epsilon|\epsilon > c]$ for any constant c will exceed $E[\epsilon]$.

Censored and Truncated Means in Linear Regression

- For data left-censored at zero suppose we observe $y = 0$, rather than merely that $y^* \leq 0$.
- The censored mean is obtained by first conditioning the observable y on the binary indicator d defined in (1) with $L = 0$ and then unconditioning.
- Suppressing dependence on x for notational simplicity again, we have the left-censored mean

$$\begin{aligned} E[y] &= E_d[E_{y|d}[y|d]] \\ &= Pr[d = 0] \times E[y|d = 0] + Pr[d = 1] \times E[y|d = 1] \\ &= 0 \times Pr[y^* \leq 0] + Pr[y^* > 0] \times E[y^*|y^* > 0] \\ &= Pr[y^* > 0] \times E[y^*|y^* > 0] \end{aligned}$$

Censored and Truncated Means in Linear Regression

- Remember $Pr[y^* > 0] = 1 - Pr[y^* \leq 0] = Pr[\epsilon > -x'\beta]$ and $E[y^*|y^* > 0]$ is given by (10).
- Summarizing, for the linear regression model with censoring or truncation from below at zero, the conditional means are given by

$$\begin{aligned}
\text{latent variable:} \quad & E[y^*|x] = x'\beta && (11)\\
\text{left truncated at 0:} \quad & E[y|x, y > 0] = x'\beta + E[\epsilon|\epsilon > -x'\beta] && (12)\\
\text{left censored at 0:} \quad & E[y|x] = Pr[\epsilon > -x'\beta]\{x'\beta + E[\epsilon|\epsilon > -x'\beta]\} && (13)
\end{aligned}$$

Censored and Truncated Means in Linear Regression

- It is clear that even though the original conditional mean is linear, censoring or truncation leads to conditional means that are nonlinear so that OLS estimates will be inconsistent.
- One possible approach to take is a parametric one of assuming a distribution for ϵ. This leads to expressions for $E[\epsilon|\epsilon > -x'\beta]$ and $Pr[\epsilon > -x'\beta]$ and hence the truncated or censored conditional mean.

Censored and Truncated Means in the Tobit Model

- For the Tobit model the regression error ϵ is normal
- **Truncated Moments of the Standard Normal:** Suppose $z \sim N[0,1]$. Then the left-truncated moments of z are
  - $E[z|z > c] = \phi(c)/[1 - \Phi(c)]$, and $E[z|z > -c] = \phi(c)/\Phi(c)$,
  - $E[z^2|z > c] = 1 + c\phi(c)/[1 - \Phi(c)]$, and
  - $V[z|z > c] = 1 + c\phi(c)/[1 - \Phi(c)] - \phi(c)^2/[1 - \Phi(c)]^2$

Censored and Truncated Means in the Tobit Model

- Appling this result to (10), the error term has truncated mean

$$\begin{aligned} E[\epsilon|\epsilon > -x'\beta] &= \sigma E[\frac{\epsilon}{\sigma}|\frac{\epsilon}{\sigma} > \frac{-x'\beta}{\sigma}] \\ &= \sigma\phi(\frac{x'\beta}{\sigma})/[\Phi(\frac{x'\beta}{\sigma})] \\ &= \sigma\lambda(\frac{x'\beta}{\sigma}) \qquad (14) \end{aligned}$$

where $\lambda(z) = \phi(z)/\Phi(z)$ is the inverse Mills ratio.

Censored and Truncated Means in the Tobit Model

- Then the conditional means in (11)-(13) specialize to

$$\text{latent variable:} \quad E[y^*|x] = x'\beta \tag{15}$$

$$\text{left truncated at 0:} \quad E[y|x, y > 0] = x'\beta + \sigma\lambda(\frac{x'\beta}{\sigma}) \tag{16}$$

$$\text{left censored at 0:} \quad E[y|x] = \Phi(\frac{x'\beta}{\sigma})\{x'\beta + \sigma\phi(\frac{x'\beta}{\sigma})\} \tag{17}$$

Censored and Truncated Means in the Tobit Model

- The variance is similarly obtained. Defining $w = x'\beta/\sigma$, we have

$$\begin{aligned} \text{latent variable:} \quad & V[y^*|x] = \sigma^2 \qquad (18) \\ \text{left truncated at 0:} \quad & V[y|x, y > 0] = \sigma^2[1 - w\lambda(w) - \lambda(w)^2] \\ \text{left censored at 0:} \quad & V[y|x] = \sigma^2\Phi(w)\{w^2 + w\lambda(w) + \\ & \qquad +1 - \Phi(w)[w + \lambda(w)]\}^2 \\ & \qquad (19) \end{aligned}$$

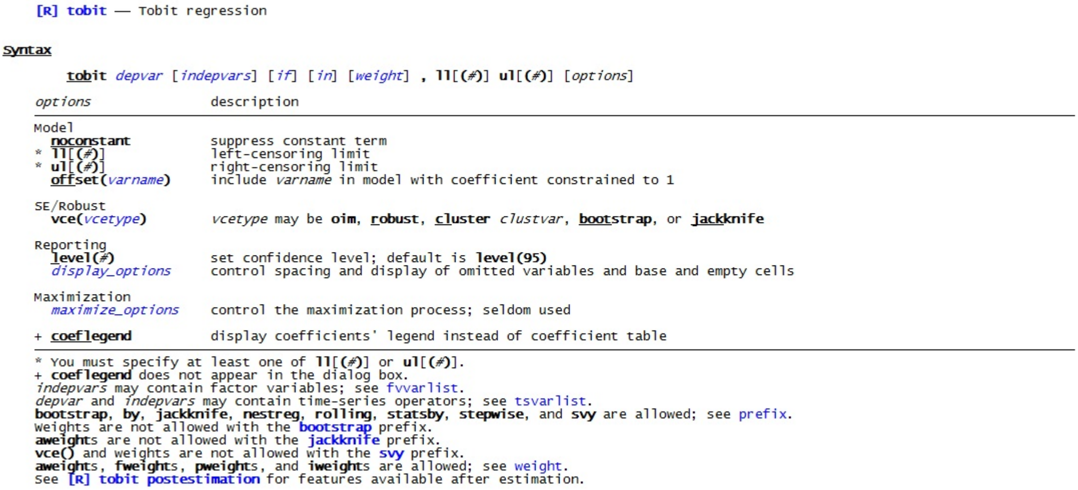
- Clearly truncation and censoring induce heteroskedasticity, and for truncation $V[y|x] < \sigma^2$ so that truncation reduces variability, as expected.

Marginal Effects in the Tobit Model

- The marginal effect is the effect on the conditional mean of the dependent variable of changes in the regressors.
- This effect varies according to whether interest lies in the latent variable mean $x'\beta$ or the truncated or censored means given in (15)-(17).

$$\begin{aligned} \text{latent variable:} \quad & \partial E[y^*|x]/\partial x = \beta & (20) \\ \text{left truncated at 0:} \quad & \partial E[y|x, y > 0]/\partial x = [1 - w\lambda(w) - \lambda(w)^2]\beta & \\ \text{left censored at 0:} \quad & \partial E[y|x]/\partial x = \Phi(w)\beta & (21) \end{aligned}$$

Tobit en Stata



```
    [R] tobit — Tobit regression

Syntax

        tobit depvar [indepvars] [if] [in] [weight] , ll[(#)] ul[(#)] [options]

    options                 description
    ---------------------------------------------------------------------------------------------
    Model
      noconstant            suppress constant term
    * ll[(#)]               left-censoring limit
    * ul[(#)]               right-censoring limit
      offset(varname)       include varname in model with coefficient constrained to 1

    SE/Robust
      vce(vcetype)          vcetype may be oim, robust, cluster clustvar, bootstrap, or jackknife

    Reporting
      level(#)              set confidence level; default is level(95)
      display_options       control spacing and display of omitted variables and base and empty cells

    Maximization
      maximize_options      control the maximization process; seldom used

    + coeflegend            display coefficients' legend instead of coefficient table
    ---------------------------------------------------------------------------------------------
    * You must specify at least one of ll[(#)] or ul[(#)].
    + coeflegend does not appear in the dialog box.
    indepvars may contain factor variables; see fvvarlist.
    depvar and indepvars may contain time-series operators; see tsvarlist.
    bootstrap, by, jackknife, nestreg, rolling, statsby, stepwise, and svy are allowed; see prefix.
    Weights are not allowed with the bootstrap prefix.
    aweights are not allowed with the jackknife prefix.
    vce() and weights are not allowed with the svy prefix.
    aweights, fweights, pweights, and iweights are allowed; see weight.
    See [R] tobit postestimation for features available after estimation.
```

Sample Selection Models

- Selection may be due to self-selection, with the outcome of interest determined in part by individual choice of whether or not to participate in the activity of interest.
- It can also result from sample selection, with those who participate in the activity of interest deliberately oversampled - an extreme case being sampling only participants.
- In either case, similar issues arise and selection models are usually called sample selection models.

A Bivariate Sample Selection Model (Type II Tobit)

- Let y_1^* denote the outcome of interest. In the standard truncated Tobit model this outcome is observed if $y_1^* > 0$.
- A more general model introduces a different latent variable, y_2^*, and the outcome y_1^* is observed if $y_2^* > 0$.
- For example, y_2^* determines whether or not to work and y_1^* determines how much to work, and $y_1^* \neq y_2^*$ since there are fixed costs to work such as commuting costs that are more important in determining participation than hours of work once working.

A Bivariate Sample Selection Model (Type II Tobit)

- The bivariate sample selection model comprises a participation equation that

$$y_2 = \begin{cases} 1 \text{ if } y_2^* > 0, \\ 0 \text{ if } y_2^* \leq 0 \end{cases} \tag{22}$$

- and a resultant outcome equation that

$$y_1 = \begin{cases} y_1^* \text{ if } y_2^* > 0, \\ - \text{ if } y_2^* \leq 0 \end{cases} \tag{23}$$

- This model specifies that y_1 is observed when $y_2^* > 0$, whereas y_1 need not take on any meaningful value when $y_2^* \leq 0$.

A Bivariate Sample Selection Model (Type II Tobit)

- In general the model can be written for a random draw from the population as

$$\begin{aligned} y_1 &= x_1\beta_1 + u_1 && (24) \\ y_2 &= 1[x\delta_2 + v_2 > 0] && (25) \end{aligned}$$

- **Assumption:** (a) (x, y_2) is always observed in the population, but y_1 is observed only when $y_2 = 1$; (b) $(u_1; v_2)$ is independent of x with zero mean; (c) $v_2 \sim Normal(0; 1)$; and (d) $E(u_1|v_2) = \gamma_1 v_2$.
- Assumption (d) requires linearity in the population regression of u_1 on v_2. It always holds if $(u_1; v_2)$ is bivariate normal.

A Bivariate Sample Selection Model (Type II Tobit)

- To derive an estimating equation, let $(y_1; y_2; x; u_1; v_2)$ denote a random draw from the population.
- Since y_1 is observed only when $y_2 = 1$, what we can hope to estimate is $E(y_1|x; y2 = 1)$ [along with $P(y_2 = 1|x)$].
- From (24)

$$E(y_1|x; v_2) = x_1\beta_1 + E(u_1|x; v_2) = x_1\beta_1 + E(u_1|v_2) = x_1\beta_1 + \gamma_1 v_2 \tag{26}$$

- If $\gamma_1 = 0$ (which implies that u_1 and v_2 are uncorrelated) then $E(y_1|x; v_2) = E(y_1|x) = E(y_1|x_1) = x_1\beta_1$.
- In other words, if $\gamma_1 = 0$, then **there is no sample selection problem**, and β_1 can be consistently estimated by OLS using the selected sample.

A Bivariate Sample Selection Model (Type II Tobit)

- If $\gamma_1 \neq 0$, then using iterated expectations in (26)

$$E(y_1|x;y_2) = x_1\beta_1 + \gamma_1 E(v_2|x;y_2) = x_1\beta_1 + \gamma_1 h(x;v_2) \tag{27}$$

  where $h(x;v_2) = E(v_2|x;y_2)$.

- Since the selected sample has $y_2 = 1$, we need only find $h(x;1)$. But $h(x;1) = E(v_2|v_2 > -x\delta_2) = \lambda(x\delta_2)$, where $\lambda(\cdot) \equiv \phi(\cdot)/\Phi(\cdot)$ is the inverse Mills ratio.
- We can write

$$E(y_1|x;y_2 = 1) = x_1\beta_1 + \gamma_1\lambda(x\delta_2) \tag{28}$$

A Bivariate Sample Selection Model (Type II Tobit)

- Equation (28) makes it clear that an OLS regression of y_1 on x_1 using the selected sample omits the term $\lambda(x\delta_2)$ and generally leads to inconsistent estimation of β_1.
- Equation (28) also suggests a way to consistently estimate β_1.
- Following Heckman (1979), we can consistently estimate β_1 and γ_1 using the selected sample by regressing y_1 on x_1, and $\lambda(x\delta_2)$.
- The problem is that δ_2 is unknown, so we cannot compute the additional regressor $\lambda(x\delta_2)$.
- However, a consistent estimator of δ_2 is available from the first-stage probit estimation of the selection equation.

A Bivariate Sample Selection Model (Type II Tobit)

- Heckman's Procedure (Heckit estimator): (a) Obtain the probit estimate $\hat{\delta}_2$ from the model
$$P(y_{i2} = 1|x_i) = \Phi(x_i\delta_2) \tag{29}$$
using **all N observations.** Then obtain the estimated inverse Mills ratios $\hat{\lambda}_{i2} \equiv \lambda(x_i\hat{\delta}_2)$ (at least for $i = 1, \ldots, N_1$).
(b) Obtain $\hat{\beta}_1$ and $\hat{\gamma}_1$ from the OLS regression on the selected sample, y_{i1} on x_{i1}; and $\hat{\lambda}_{i2}$; for $i = 1, \ldots, N_1$.
- These estimators are consistent and $\sqrt{N}$ asymptotically normal.

A Bivariate Sample Selection Model (Type II Tobit)

- When $\gamma_1 \neq 0$, obtaining a consistent estimate for the asymptotic variance of $\hat{\beta}_1$ is complicated for two reasons.

1. If $\gamma_1 \neq 0$, then $Var(y1|x; y_2 = 1)$ is not constant. As we know, heteroskedasticity itself is easy to correct for using the robust standard errors. However,
2. we should also account for the fact that $\hat{\delta}_2$ is an estimator of δ_2.

- The adjustment to the variance of $(\hat{\beta}_1; \hat{\gamma}_1)$ because of the two-step estimation is cumbersome, it is not enough to simply make the standard errors heteroskedasticity robust.

A Bivariate Sample Selection Model (Type II Tobit)

- Replacing parts (c) and (d) in the Assumption above with the stronger assumption that $(u_1; v_2)$ is bivariate normal with mean zero, $Var(u_1) = \sigma_1^2$, $Cov(u_1; v_2) = \sigma_{12}$, and $Var(v_2) = 1$, then partial maximum likelihood estimation can be used.

$$\begin{bmatrix} v_2 \\ u_1 \end{bmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sigma_{12} \\ \sigma_{12} & \sigma_1^2 \end{pmatrix}\right] \tag{30}$$

- Given participation and outcome equations, for $y_2^* > 0$ we observe y_1, with probability equal to the probability that $y_2^* > 0$ times the conditional probability of y_1^* given that $y_2^* > 0$.
- For positive y_1 the density of observables is $f^*(y_1^*|y_2^* > 0) \times Pr[y_2^* > 0]$.
- For $y_2^* \leq 0$ all that is observed is that this event has occurred, and the density is the probability of this event occurring.

A Bivariate Sample Selection Model (Type II Tobit)

- The bivariate sample selection model therefore has likelihood function

$$L = \prod_{i=1}^{n} [Pr(y_{2i}^* \leq 0)]^{1-y_{2i}} \{f(y_{1i}|y_{2i}^* > 0) \times Pr[y_{2i}^* > 0]\}^{y_{2i}} \tag{31}$$

  where the first term is the discrete contribution when $y_{2i}^* \leq 0$, since then $y_{2i} = 0$, and the second term is the continuous contribution when $y_2^* i > 0$.

- The classic early application of this model was to labor supply, where y_2 is the unobserved desire or propensity to work, whereas y_1 is actual hours worked.

  ▸ Go to Example

Sample Selection: Type III Tobit Model

- Consider the case where the selection equation is of the censored Tobit form.
- The population model is

$$\begin{aligned} y_1 &= x_1\beta_1 + u_1 \qquad (32) \\ y_2 &= \max(0; x\delta_2 + v_2) \qquad (33) \end{aligned}$$

  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

- The starting point is equation (26), just as in the probit selection case.
- 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

$$E(y_1 | x; v_2; s_2 = 1) = x_1\beta_1 + \gamma_1 v_2 \tag{34}$$

- 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 (33) by standard Tobit using all N observations. For $y_{i2} > 0$ (say $i = 1, 2, \ldots N_1$), define

$$\hat{v}_{i2} = y_{i2} - xi\hat{\delta}_2 \tag{35}$$

  (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 $(\beta_1; \gamma_1)$.

Sesgo de Selección por Truncamiento Incidental

- En economía un caso emblemático de aplicación de máxima verosimilitud 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, número de hijos etc.
- La segunda ecuación es una ecuación de horas deseadas de trabajo que depende del salario, de la presencia de hijos pequeños en el hogar, del estado civil, etc.

Sesgo de Selección por Truncamiento Incidental

- 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.
- Definiciones: 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:

$$f(y, z|z > a) = \frac{f(y, z)}{Pr(z > a)}$$

Sesgo de Selección por Truncamiento Incidental

- Teorema 20.4 (Greene, 1997, Cap. 20, página 975): Si y y z tienen una distribución normal bivariada con medias μ_y y μ_z, desvíos estándar σ_y y σ_z y correlación ρ, entonces:

$$E(y|z > a) = \mu_y + \rho\sigma_y\lambda(\alpha_z)$$

$$Var(y|z > a) = \sigma_y^2[1 - \rho^2\delta(\alpha_z)],$$

donde, $\alpha_z = (a - \mu_z)/\sigma_z$, $\lambda(\alpha_z) = \phi(\alpha_z)/[1 - \Phi(\alpha_z)]$ y $\delta(\alpha_z) = \lambda(\alpha_z)[\lambda(\alpha_z) - \alpha_z]$.

Sesgo de Selección por Truncamiento Incidental

- Para poner el ejemplo de la oferta de trabajo de las mujeres en un marco general de análisis, digamos que la ecuación que determina la selección muestral es:

$$z_i^* = \gamma' w_i + u_i,$$

  donde z_i^* es la diferencia entre el salario de mercado y el salario de reserva de la persona i.

- La ecuación de interés es,

$$y_i = \beta' x_i + \epsilon_i,$$

  donde y_i es la oferta de trabajo (en horas) de la persona i.

- La regla es que y_i es observada solo cuando z_i^* es mayor a cero.

Sesgo de Selección por Truncamiento Incidental

- Asumamos que u_i y ϵ_i tienen distribución normal bivariada con medias cero y correlación ρ. Aplicando el teorema 20.4 tenemos,

$$\begin{aligned}E(y_i|y_i\text{es observada}) &= E(y_i|z_i^* > 0)\\ &= E(y_i|u_i > -\gamma' w_i)\\ &= \beta' x_i + E(\epsilon_i|u_i > -\gamma' w_i)\\ &= \beta' x_i + \rho\sigma_\epsilon\lambda_i(\alpha_u)\\ &= \beta' x_i + \beta_\lambda\lambda_i(\alpha_u)\end{aligned}$$

  donde $\alpha_u = -\gamma' w_i/\sigma_u$ y $\lambda_i(\alpha_u) = \phi(\gamma' w_i/\sigma_u)/\Phi(\gamma' w_i/\sigma_u)$.

- Entonces, la ecuación de interés puede escribirse como,

$$y_i|z_i^* > 0 = \beta' x_i + \beta_\lambda\lambda_i(\alpha_u) + v_i$$

  donde v_i es un término de error con media cero.

Sesgo de Selección por Truncamiento Incidental

- Como queda claro de este desarrollo, estimar por MCC la ecuación de interés usando solo los datos observados, produce estimadores inconsistentes de β por el argumento estándar de variables omitidas (i.e. estamos omitiendo $\lambda_i(\cdot)$).
- Cómo podemos obtener estimaciones consistentes de la ecuación de horas de trabajo utilizando solo los datos observados.
- En principio tenemos un problema similar al de la variable habilidad omitida en la ecuación del salario. En este caso, la variable $\lambda_i(\cdot)$ no es observada.
- Note que aún cuando observáramos $\lambda_i(\cdot)$, MCC no nos daría estimadores eficientes porque los errores de la ecuación de interés, v_i, son heterocedásticos (i.e. $Var(v_i) = \sigma_\epsilon^2(1 - \rho^2\delta_i)$ de acuerdo al teorema 20.4 de Greene).

Sesgo de Selección por Truncamiento Incidental

- Una posible solución es estimar la ecuación de selección para obtener los $\hat{\gamma}$ y construir la variable omitida como $\hat{\lambda}_i = \phi(\hat{\gamma}' w_i)/\Phi(\hat{\gamma}' w_i)$. Luego en un segundo paso estimar la ecuación de interés por MCC en una regresión de y sobre x y $\hat{\lambda}$.
- El único problema de esta solución es que la variable dependiente de la ecuación de selección, z_i^*, no es observada. Lo que podemos observar es $z_i = 1$ si $z_i^* > 0$, es decir si la persona está trabajando; o $z_i = 0$ si $z_i^* < 0$ si la persona no está trabajando.
- Es decir que el modelo que podemos estimar es:

$$z_i = \gamma' w_i + u_i, \tag{36}$$

donde z_i es una variable binaria.

Sesgo de Selección por Truncamiento Incidental

- Heckman (1979) sugirió utilizar el siguiente procedimiento en dos etapas.
  1. Estimar la ecuación de selección usando un Probit para obtener estimaciones de γ. Luego para cada observación de la muestra se construye,

     $$\hat{\lambda}_i = \frac{\phi(\hat{\gamma}' w_i)}{\Phi(\hat{\gamma}' w_i)}$$

     También vamos a necesitar $\hat{\delta}_i = \hat{\lambda}_i(\hat{\lambda}_i + \hat{\gamma}' w_i)$.
  2. Estime β y β_λ por MCC en una regresión de y sobre x y $\hat{\lambda}$.

Sesgo de Selección por Truncamiento Incidental

- Para poder hacer inferencia estadística en la regresión del segundo paso hay que tener en cuenta dos problemas: heterocedasticidad en v_i y el hecho de que una de las variables explicativas de la regresión está construída a partir de una estimación anterior.
- Heckman (1979) deriva la verdadera matriz de varianzas y covarianzas de los estimadores del segundo paso.
- Recuerde que $\widehat{Var(v_i)} = \hat{\sigma}_\epsilon^2(1 - \hat{\rho}^2\hat{\delta}_i)$ usando el teorema (20.4) de Greene. Donde $\hat{\sigma}_\epsilon^2 = \frac{e'e}{n} + \bar{\hat{\delta}}\hat{\beta}_\lambda^2$, y $\bar{\hat{\delta}} = \frac{1}{n}\sum_i \hat{\delta}_i$.
- Entonces la estimación correcta de la matriz de varianzas y covarianzas del segundo paso es,

$$Var[\hat{\beta},\, \hat{\beta}_\lambda] = \hat{\sigma}_\epsilon^2(\sum_i x_i^{*'} x_i^*)^{-1}\left[\sum_i (1 - \hat{\rho}^2\hat{\delta}_i)x_i^* x_i^{*'} + Q\right](\sum_i x_i^{*'} x_i^*)^{-1}$$

Sesgo de Selección por Truncamiento Incidental

- Donde,

$$Q = \hat{\rho}^2 (X^{*'} \Delta W)[Avar(\hat{\gamma})](W' \Delta X^*)$$

y Δ es una matriz diagonal con $\hat{\delta}_i$ en la diagonal principal.

▸ Return to Sample Selection