Tobit and Selection Models

Class Notes

Manuel Arellano

November 24, 2008

1 Censored Regression

1.1 Illustration 1: Top-coding in wages

Suppose Y (log wages) are subject to "top coding" (as is often the case with social security records):

$$Y = \begin{cases} Y^* \text{ if } Y^* \leq c \\ c \text{ if } Y^* > c \end{cases}$$

Suppose we are interested in $E(Y^*)$. Effectively it is not identified but if we assume $Y^* \sim \mathcal{N}(\mu, \sigma^2)$, then μ can be determined from the distribution of Y.

The density of Y is of the form

$$f(r) = \begin{cases} \frac{1}{\sigma}\phi\left(\frac{r-\mu}{\sigma}\right) \text{ if } r < c \\ \Pr(Y^* \geq c) = 1 - \Phi\left(\frac{r-\mu}{\sigma}\right) \text{ if } r \geq c \end{cases}$$

The log-likelihood function of the sample $\{y_1, ..., y_N\}$ is

$$\mathcal{L}(\mu, \sigma^2) = \prod_{y_i<c} \frac{1}{\sigma}\phi\left(\frac{y_i - \mu}{\sigma}\right) \prod_{y_i=c}\left[1 - \Phi\left(\frac{c-\mu}{\sigma}\right)\right].$$

Usually, we shall be interested in a regression version of this model:

$$Y^* \mid X = x \sim \mathcal{N}\left(x'\beta, \sigma^2\right),$$

in which case the likelihood is of the form

$$\mathcal{L}(\beta, \sigma^2) = \prod_{y_i<c} \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right) \prod_{y_i=c}\left[1 - \Phi\left(\frac{c - x'\beta}{\sigma}\right)\right].$$

Means of censored normal variables   Consider the following right-censored variable:

$$Y = \begin{cases} Y^* \text{ if } Y^* \leq c \\ c \text{ if } Y^* > c \end{cases}$$

with $Y^* \sim \mathcal{N}(\mu, \sigma^2)$. Therefore,

$$E(Y) = E(Y^* \mid Y^* \leq c)\Pr(Y^* \leq c) + c\Pr(Y^* > c)$$

Letting $Y^* = \mu + \sigma\varepsilon$ with $\varepsilon \sim \mathcal{N}(0,1)$

$$\Pr(Y^* \leq c) = \Phi\left(\frac{c-\mu}{\sigma}\right)$$

$$E(Y^* \mid Y^* \leq c) = \mu + \sigma E\left(\varepsilon \mid \varepsilon \leq \frac{c-\mu}{\sigma}\right) = \mu - \sigma\lambda\left(\frac{c-\mu}{\sigma}\right).$$

Note that

$$E(\varepsilon \mid \varepsilon \leq r) = \int_{-\infty}^{r} e\frac{\phi(e)}{\Phi(r)}de = -\frac{1}{\Phi(r)}\int_{-\infty}^{r}\phi'(e)\,de = -\frac{\phi(r)}{\Phi(r)} = -\lambda(r) \tag{1}$$

and

$$E(\varepsilon \mid \varepsilon > r) = \int_{r}^{\infty} e\frac{\phi(e)}{\Phi(-r)}de = -\frac{1}{\Phi(-r)}\int_{r}^{\infty}\phi'(e)\,de = -\frac{-\phi(r)}{\Phi(-r)} = \frac{\phi(-r)}{\Phi(-r)} = \lambda(-r). \tag{2}$$

1.2   Illustration 2: Censoring at zero (Tobit model)

Tobin (1958) considered the following model for expenditure on durables

$$Y = \max\left(X'\beta + U, 0\right)$$
$$U \mid X \sim \mathcal{N}\left(0, \sigma^2\right).$$

This is similar to the first example, but now we have left-censoring at zero. However, the nature of the application is very different because there is no physical censoring (the variable Y^* is just a model's construct). We are interested in the model as a way of capturing a particular form of nonlinearity in the relationship between X and Y. In a utility based model, the variable Y^* might be interpretable as a notional demand before non-negativity is imposed.

With censoring at zero we have

$$Y = \begin{cases} Y^* \text{ if } Y^* > 0 \\ 0 \text{ if } Y^* \leq 0 \end{cases}$$

$$E(Y) = E(Y^* \mid Y^* > 0)\Pr(Y^* > 0)$$

$$\Pr(Y^* > 0) = \Pr\left(\varepsilon > -\frac{\mu}{\sigma}\right) = \Phi\left(\frac{\mu}{\sigma}\right)$$

$$E(Y^* \mid Y^* > 0) = \mu + \sigma E\left(\varepsilon \mid \varepsilon > -\frac{\mu}{\sigma}\right) = \mu + \sigma\lambda\left(\frac{\mu}{\sigma}\right).$$

2 Heckman's generalized selection model

Consider the model

$$y^* = x'\beta + \sigma u$$
$$d = 1(z'\gamma + v \geq 0)$$
$$\begin{pmatrix} u \\ v \end{pmatrix} \mid z \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$$

so that

$$v \mid z, u \sim \mathcal{N}\left(\rho u, 1 - \rho^2\right) \qquad \text{or } \Pr\left(v \leq r \mid z, u\right) = \Phi\left(\frac{r - \rho u}{\sqrt{1 - \rho^2}}\right).$$

In Heckman's original model, y^* denotes female log market wage and d is an indicator of participation in the labor force. The index $\{z'\gamma + v\}$ is a reduced form of the difference between market wage and reservation wage.

Joint likelihood function   The joint likelihood is:

$$L = \sum_{d=1} \ln\left\{p\left(d = 1, y^* \mid z\right)\right\} + \sum_{d=0} \ln \Pr\left(d = 0 \mid z\right)$$

we have

$$p\left(d = 1, y^* \mid z\right) = \Pr\left(d = 1 \mid z, y^*\right) f\left(y^* \mid z\right)$$

$$f\left(y^* \mid z\right) = \frac{1}{\sigma}\phi\left(\frac{y^* - x'\beta}{\sigma}\right)$$

$$\Pr\left(d = 1 \mid z, y^*\right) = 1 - \Pr\left(v \leq -z'\gamma \mid z, u\right) = 1 - \Phi\left(\frac{-z'\gamma - \rho u}{\sqrt{1 - \rho^2}}\right) = \Phi\left(\frac{z'\gamma + \rho u}{\sqrt{1 - \rho^2}}\right).$$

Thus

$$L\left(\gamma, \beta, \sigma\right) = \sum_{d=1}\left\{\ln\left[\frac{1}{\sigma}\phi\left(u\right)\right] + \ln \Phi\left(\frac{z'\gamma + \rho u}{\sqrt{1 - \rho^2}}\right)\right\} + \sum_{d=0} \ln\left[1 - \Phi\left(z'\gamma\right)\right]$$

where

$$u = \frac{y^* - x'\beta}{\sigma}.$$

Note that if $\rho = 0$ this log likelihood boils down to the sum a Gaussian linear regression log likelihood and a probit log likelihood.

Density of y^* conditioned on $d = 1$ From the previous result we know that

$$p\left(d = 1, y^* \mid z\right) = \frac{1}{\sigma}\phi\left(\frac{y^* - x'\beta}{\sigma}\right)\Phi\left(\frac{z'\gamma + \rho u}{\sqrt{1 - \rho^2}}\right).$$

Alternatively, to obtain it we could factorize as follows

$$p\left(d = 1, y^* \mid z\right) = \Pr\left(d = 1 \mid z\right) f\left(y^* \mid z, d = 1\right) = \Phi\left(z'\gamma\right) f\left(y^* \mid z, d = 1\right).$$

From the previous expression we know that

$$f\left(y^* \mid z, d = 1\right) = \frac{p\left(d = 1, y^* \mid z\right)}{\Phi\left(z'\gamma\right)} = \frac{1}{\Phi\left(z'\gamma\right)}\Phi\left(\frac{z'\gamma + \rho u}{\sqrt{1 - \rho^2}}\right)\frac{1}{\sigma}\phi\left(u\right).$$

Note that if $\rho = 0$ we have $f\left(y^* \mid z, d = 1\right) = f\left(y^* \mid z\right) = \sigma^{-1}\phi\left(u\right)$.

Two-step method Then mean of $f\left(y^* \mid z, d = 1\right)$ is given by

$$\begin{aligned} E\left(y^* \mid z, d = 1\right) &= x'\beta + \sigma E\left(u \mid z'\gamma + v \geq 0\right) \\ &= x'\beta + \sigma\rho E\left(v \mid v \geq -z'\gamma\right) = x'\beta + \sigma\rho\lambda\left(z'\gamma\right) \end{aligned}$$

Form $w_i = \left(x_i', \widehat{\lambda}_i\right)'$, where $\widehat{\lambda}_i = \lambda\left(z_i'\widehat{\gamma}\right)$ and $\widehat{\gamma}$ is the probit estimate. Then do the OLS regression of y on x and $\widehat{\lambda}$ in the subsample with $d = 1$ to get consistent estimates of β and $\sigma_{uv}\left(= \sigma\rho\right)$:

$$\begin{pmatrix} \widehat{\beta} \\ \widehat{\sigma}_{uv} \end{pmatrix} = \left(\sum_{d_i = 1} w_i w_i'\right)^{-1}\sum_{d_i = 1} w_i y_i.$$

2.1 Nonparametric identification: The fundamental role of exclusion restrictions

- The role of exclusion restrictions for identification in a selection model is paramount. In applications there is a marked contrast in credibility between estimates that rely exclusively on the nonlinearity and those that use exclusion restrictions.

- The model of interest is

$$\begin{aligned} Y &= g_0\left(X\right) + U \\ D &= 1\left(p\left(X, Z\right) - V > 0\right) \end{aligned}$$

where (U, V) are independent of (X, Z) and V is uniform in the $(0, 1)$ interval. Thus,

$$E\left(U \mid X, Z, D = 1\right) = E\left[U \mid V < p\left(X, Z\right)\right] = \lambda_0\left[p\left(X, Z\right)\right]$$

$$E\left(Y \mid X, Z\right) = g_0\left(X\right)$$

4

(i.e. enforcing the exclusion restriction), but we observe

$$
\begin{aligned}
E\left(Y \mid X, Z, D=1\right) &= \mu\left(X, Z\right) = g_0\left(X\right) + \lambda_0\left[p\left(X, Z\right)\right] \\
E\left(D \mid X, Z\right) &= p\left(X, Z\right).
\end{aligned}
$$

The question is whether $g_0\left(.\right)$ and $\lambda_0\left(.\right)$ are identified from knowledge of $\mu\left(X, Z\right)$ and $p\left(X, Z\right)$.

- Let us consider first the case where X and Z are continuous. Suppose there is an alternative solution (g^*, λ^*). Then

$$
g_0\left(X\right) - g^*\left(X\right) + \lambda_0\left(p\right) - \lambda^*\left(p\right) = 0.
$$

Differentiating

$$
\begin{aligned}
\frac{\partial\left(\lambda_0 - \lambda^*\right)}{\partial p}\frac{\partial p}{\partial Z} &= 0 \\
\frac{\partial\left(g_0 - g^*\right)}{\partial X} + \frac{\partial\left(\lambda_0 - \lambda^*\right)}{\partial p}\frac{\partial p}{\partial X} &= 0.
\end{aligned}
$$

Under the assumption that $\partial p / \partial Z \neq 0$ (instrument relevance), we have

$$
\frac{\partial\left(\lambda_0 - \lambda^*\right)}{\partial p} = 0, \qquad \frac{\partial\left(g_0 - g^*\right)}{\partial X} = 0
$$

so that $\lambda_0 - \lambda^*$ and $g_0 - g^*$ are constant (i.e. $g_0\left(X\right)$ is identified up to an unknown constant). This is the identification result of Das, Newey, and Vella (2003).

- $E\left(Y \mid X\right)$ is identified up to a constant, provided we have a continuous instrument. Identification of the constant requires units for which the probability of selection is arbitrarily close to one ("identification at infinity"). Unfortunately, the constants are important for identifying average treatment effects.

- With binary Z, functional form assumptions play a more fundamental role in securing identification than in the case of an exclusion restriction of a continuous variable.

- Suppose X is continuous but Z is a dummy variable. In general $g_0\left(X\right)$ is not identified. To see this, consider

$$
\begin{aligned}
\mu\left(X, 1\right) &= g_0\left(X\right) + \lambda_0\left[p\left(X, 1\right)\right] \\
\mu\left(X, 0\right) &= g_0\left(X\right) + \lambda_0\left[p\left(X, 0\right)\right],
\end{aligned}
$$

so that we identify the difference

$$
\nu\left(X\right) = \lambda_0\left[p\left(X, 1\right)\right] - \lambda_0\left[p\left(X, 0\right)\right],
$$

but this does not suffice to determine λ_0 up to a constant. Take as an example the case where $p(X, Z)$ is a simple logit or probit model of the form

$$p(X, Z) = F(\beta X + \gamma Z),$$

then letting $h_0(.) = \lambda_0[F(.)]$,

$$\nu(X) = h_0(\beta X + \gamma) - h_0(\beta X).$$

Suppose the existence of another solution h^*. We should have

$$h_0(\beta X + \gamma) - h^*(\beta X + \gamma) = h_0(\beta X) - h^*(\beta X),$$

which is satisfied by a multiplicity of periodic functions.

- If X is also discrete, there is clearly lack of identification. For example, suppose X and Z are dummy variables:

$$
\begin{aligned}
\mu(0, 0) &= g_0(0) + \lambda_0[p(0, 0)] \\
\mu(0, 1) &= g_0(0) + \lambda_0[p(0, 1)] \\
\mu(1, 0) &= g_0(1) + \lambda_0[p(1, 0)] \\
\mu(1, 1) &= g_0(1) + \lambda_0[p(1, 1)].
\end{aligned}
$$

Since $\lambda_0(.)$ is unknown $g_0(1) - g_0(0)$ is not identified. Only $\lambda_0[p(1, 1)] - \lambda_0[p(1, 0)]$ and $\lambda_0[p(0, 1)] - \lambda_0[p(0, 0)]$ are identified.

3   Relaxing restrictive distributional assumptions in Tobit models

- Symmetrically trimmed least squares.

- Censored regression quantiles.

4   Bivariate probit with sample selection

See Exercise 8 in the 'Discrete choice, censoring, and duration' list.