

Censored Outcome Models: Tobit

Anders Munk-Nielsen, Jesper Riis-Vestergaard Sørensen

November, 2021

Corner Outcome Models

The purpose of this week's problem set is to estimate a censored regression model known as the Tobit Model. Censored regression models are relevant when the dependent variable has positive probability mass at some point, but is continuous in other ranges. One example is top-coded income data, where the only available information is whether a particular individual has income at or above some threshold level. In this case, the income data are right-censored and the censoring is basically a "flaw" in the data. A variable can also be left-censored. In this case, the observed variable equals the threshold value when the original variable is less than or equal to the threshold level. Another situation where censored regression models can be applied is the case of corner solutions. These arise, e.g., in consumption problems, where a corner solution exists when a consumer consumes nothing of a particular commodity. In this case, the data is not "flawed" but censoring is due to the behaviour of the economic agent.

A censored variable has positive probability mass at some point. The case of a left-censored variable is illustrated in Figure 1 below.

The Tobit Model

Consider the Tobit model for a randomly drawn observation

$$y_i^* = \mathbf{x}_i\beta + u_i \quad \text{where } u_i|\mathbf{x}_i \sim \mathcal{N}(0, \sigma^2) \quad (1)$$

$$y_i = \max(0, y_i^*) \quad (2)$$

y_i^* is a latent variable that is incompletely observed: y_i is observed only when $y_i^* > 0$. \mathbf{x}_i is an observed vector of covariates and β is a vector of parameters which we seek to estimate. With respect to the latter, the crucial assumption is that the random independent error term, u_i , is normal.

The conditional expectation $E(y|\mathbf{x})$ is

Figure 1: CENSORED DISTRIBUTION

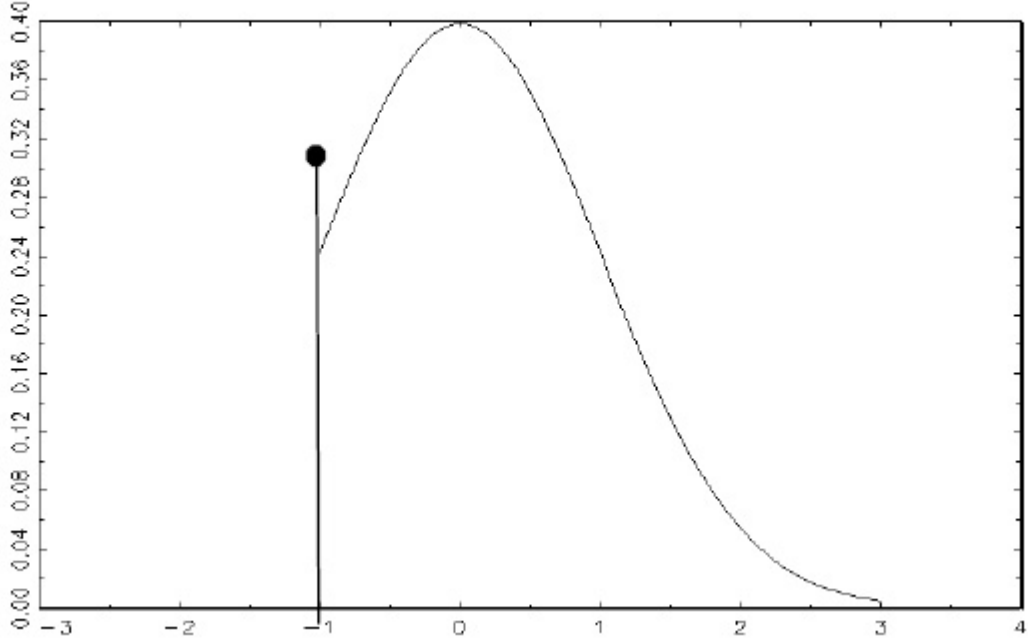


Figure 1: image

$$\begin{aligned} E(y|\mathbf{x}) &= \Pr(y = 0|\mathbf{x}) \cdot 0 + \Pr(y > 0|\mathbf{x}) \cdot E(y|\mathbf{x}, y > 0) \\ &= \Pr(y > 0|\mathbf{x}) \cdot E(y|\mathbf{x}, y > 0), \end{aligned} \tag{1}$$

where $\Pr(y > 0|\mathbf{x})$ is given by

$$\begin{aligned} \Pr(y > 0|\mathbf{x}) &= \Pr(y^* > 0|\mathbf{x}) \\ &= \Pr(\mathbf{x}\beta + u > 0) \\ &= \Pr(u > -\mathbf{x}\beta) \\ &= \Pr\left(\frac{u}{\sigma} > \frac{-\mathbf{x}\beta}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{-\mathbf{x}\beta}{\sigma}\right) = \Phi\left(\frac{\mathbf{x}\beta}{\sigma}\right), \end{aligned} \tag{2}$$

where the second-last equality follows from the normality assumption, and the last one from the symmetry of the normal distribution. To obtain an expression for $E(y|\mathbf{x}, y > 0)$ write

$$\begin{aligned}
E(y|\mathbf{x}, y > 0) &= E(y^*|\mathbf{x}, y^* > 0) \\
&= E(\mathbf{x}\beta + u|\mathbf{x}, \mathbf{x}\beta + u > 0) \\
&= \mathbf{x}\beta + E(u|\mathbf{x}, u > -\mathbf{x}\beta) \\
&= \mathbf{x}\beta + E\left(u\left|\mathbf{x}, \frac{u}{\sigma} > \frac{-\mathbf{x}\beta}{\sigma}\right.\right).
\end{aligned} \tag{3}$$

The only difficult part here is $E(u|\mathbf{x}, u/\sigma > -\mathbf{x}\beta/\sigma)$. Rewrite this as

$$E\left(u\left|\mathbf{x}, \frac{u}{\sigma} > \frac{-\mathbf{x}\beta}{\sigma}\right.\right) = \sigma E\left(\frac{u}{\sigma}\left|\frac{u}{\sigma} > \frac{-\mathbf{x}\beta}{\sigma}\right.\right) \tag{3}$$

$$= \sigma \frac{\phi(-\mathbf{x}\beta/\sigma)}{1 - \Phi(-\mathbf{x}\beta/\sigma)} \tag{4}$$

The numerator gives the density at $-\mathbf{x}\beta/\sigma$ and the denominator gives the probability that $u/\sigma > -\mathbf{x}\beta/\sigma$, i.e., it scales the density up by the probability to observe an individual with $y^* > 0$. This is illustrated in figure 2, where the density function for a standard normal is shown together with the density of a standard normal that is truncated at zero. The truncation implies that the mean of the truncated distribution is positive, as indicated by the dotted line.

Figure 2: A STANDARD NORMAL DISTRIBUTION AND A TRUNCATED NORMAL DISTRIBUTION WHERE THE TRUNCATION POINT IS AT $-x\beta$. DOTTED LINE GIVES MEAN OF THE TRUNCATED DISTRIBUTION.

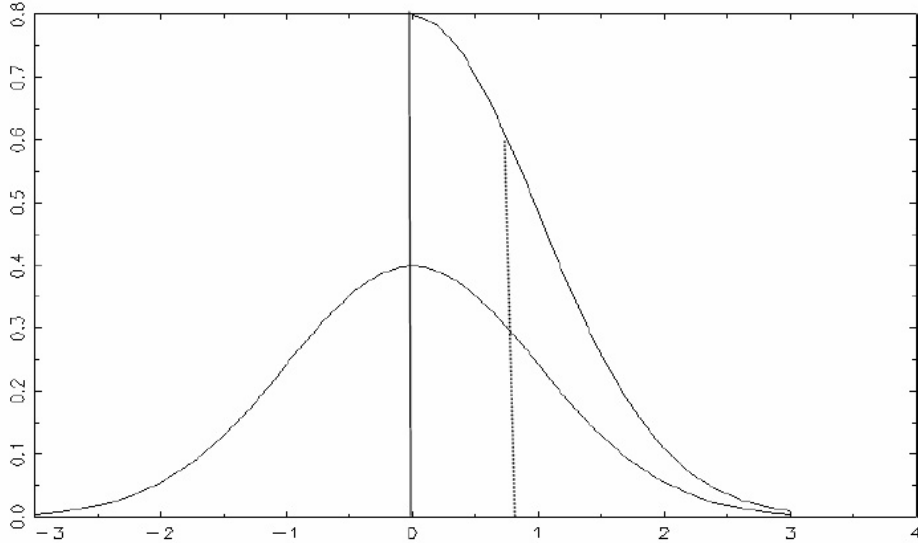


Figure 2: image

Using the symmetry of the normal distribution, $1 - \Phi(-\mathbf{x}\beta/\sigma) = \Phi(\mathbf{x}\beta/\sigma)$ and $\phi(-\mathbf{x}\beta/\sigma) = \phi(\mathbf{x}\beta/\sigma)$, we have that

$$E(u|\mathbf{x}, u > -\mathbf{x}\beta) = \sigma \frac{\phi(\mathbf{x}\beta/\sigma)}{\Phi(\mathbf{x}\beta/\sigma)}$$

The expression for $E(y|\mathbf{x}, y > 0)$ found in (3) can then be written as:

$$E(y|\mathbf{x}, y > 0) = \mathbf{x}\beta + \sigma\lambda(\mathbf{x}\beta/\sigma) \quad (4)$$

where $\lambda(\mathbf{x}\beta/\sigma) \equiv \phi(\mathbf{x}\beta/\sigma) / \Phi(\mathbf{x}\beta/\sigma)$ is called the inverse Mills ratio, while the expression in (1) becomes:

$$E(y|\mathbf{x}) = \mathbf{x}\beta \cdot \Phi\left(\frac{\mathbf{x}\beta}{\sigma}\right) + \sigma \phi\left(\frac{\mathbf{x}\beta}{\sigma}\right). \quad (5)$$

While $\lambda(z)$ does look weird, it also has sensible properties which make sense for our model. In particular, $\lambda(z) > 0$ for all $z \in \mathbb{R}$ with $\lambda(z) \rightarrow 0$ for $z \rightarrow \infty$ and $\lambda(z) \rightarrow \infty$ for $z \rightarrow -\infty$.

Estimation

The model has a distinct nonlinear component that depends on the probability that $y_i^* > 0$. The log-likelihood contribution for observation i therefore consists of two components:

$$\ell_i = \mathbf{1}\{y_i = 0\} \log[1 - \Phi(\mathbf{x}_i\beta/\sigma)] + \mathbf{1}\{y_i > 0\} \left\{ -\frac{1}{2} \left[\log(2\pi) + \log(\sigma^2) + (y_i - \mathbf{x}_i\beta)^2 / \sigma^2 \right] \right\} \quad (6)$$

For each observation, the likelihood contribution is determined either by the first term which is the probability of $y_i = 0$, or the second term which is similar to the log-likelihood of the linear regression model.

Marginal effects

If the interest lies in the marginal effect of a change in a continuous variable x_j conditional on $y > 0$ then we simply differentiate eq. (4) wrt. x_j . Note that since we will be looking at $E(y|\mathbf{x}, y > 0)$ and $E(y|\mathbf{x})$ where we condition on \mathbf{x} , this means that we will have to evaluate the expressions at the particular value of \mathbf{x} (When you see these conditional expectations, recall that you should really be thinking $E(y|\mathbf{x} = \mathbf{x}^0, y > 0)$, which will be a function of \mathbf{x}^0 . In other words, in these expressions, \mathbf{x} will be $1 \times K$ and not $N \times K$.)

In the following, we will make use of the fact that $\frac{\partial}{\partial z}\lambda(z) = -\lambda(z)[z + \lambda(z)]$. Writing $\lambda = \lambda(\mathbf{x}\beta/\sigma)$ in order to keep notation simple, the first derivative of the expression in eq. (4) provides:

$$\begin{aligned}
\frac{\partial E(y|\mathbf{x}, y > 0)}{\partial x_j} &= \beta_j + \sigma \frac{\partial \lambda}{\partial x_j} \\
&= \beta_j + \sigma [-\lambda \cdot (\mathbf{x}\beta/\sigma + \lambda)] (\beta_j/\sigma) \\
&= \beta_j [1 - \lambda \cdot (\mathbf{x}\beta/\sigma + \lambda)].
\end{aligned} \tag{7}$$

In the exercise, you will be asked to compute $\partial E(y|\mathbf{x} = \bar{\mathbf{x}}, y > 0)/\partial x_j$.

If the interest centers on $\partial E(y|\mathbf{x})/\partial x_j$, then the marginal effect is obtained by differentiating the expression in (5) with respect to \mathbf{x} :

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \beta_j \Phi\left(\frac{\mathbf{x}\beta}{\sigma}\right) + \mathbf{x}\beta \phi\left(\frac{\mathbf{x}\beta}{\sigma}\right) \frac{\beta_j}{\sigma} + \sigma \phi'\left(\frac{\mathbf{x}\beta}{\sigma}\right) \frac{\beta_j}{\sigma} \tag{8}$$

$$= \beta_j \Phi\left(\frac{\mathbf{x}\beta}{\sigma}\right), \tag{9}$$

where $\phi'(z) = \partial \phi(z)/\partial z$ denotes the first derivative of the pdf of the standard normal distribution, and has the property that $\phi'(z) = -z\phi(z)$, (Try to prove it!) thus explaining why the last two terms of eq. (8) cancel out.

Note: Both functional forms for the marginal effect takes the form of a correction multiplied with the parameter for the variable in question. For eq. (9), the marginal effect must have the same sign as β_j since $\Phi(z) > 0 \forall z$. The same is true for (7), because it turns out that the correction factor there will be strictly between zero and one.