Discrete Choice Models

Truncated and Censored Regression Models

Mohammad Arshad Rahman
(webpage: https://www.arshadrahman.com)

Course: Econometrics
Chennai Mathematical Institute

Truncated Density

The generic density of a truncated distribution is given as follows:

$$
\begin{aligned}
f(y|a < y < b) &= \frac{f(y)}{\Pr(y < b) - \Pr(y < a)} \\
&= \frac{f(y)}{\Pr(y < b)}, \qquad \text{if} \quad a = -\infty, \qquad (1) \\
&= \frac{f(y)}{1 - \Pr(y < a)}, \qquad \text{if} \quad b = \infty,
\end{aligned}
$$

where $f(y)$ is the untruncated density. We say that "y is truncated from below at a" and "y is truncated from above at b".

Truncated Normal Density

In regression, we are interested in the truncated normal distribution:

$$
\begin{aligned}
f(y|\mu,\sigma^2; a<y<b) &= \frac{\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)} \\
&= \frac{\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right)} \qquad \text{if} \quad a=-\infty, \qquad (2) \\
&= \frac{\frac{1}{\sigma}\phi\left(\frac{y-\mu}{\sigma}\right)}{1-\Phi\left(\frac{a-\mu}{\sigma}\right)} \qquad \text{if} \quad b=\infty,
\end{aligned}
$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the *pdf* and *cdf* of the standard normal distribution.

Mean & Variance

The expectation and variance of a truncated continuous random variable are given as,

$$
\begin{aligned}
E(y|a<y<b) &= \int_a^b y\, f(y|a<y<b)dy \\
V(y|a<y<b) &= \int_a^b \left[y - E(y|a<y<b)\right]^2 f(y|a<y<b)dy.
\end{aligned}
\tag{3}
$$

Mean & Variance

For the normal density, this translates into the following expressions:

$$E(y|\mu,\sigma^2; a<y<b) = \mu + \sigma \frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}$$

$$V(y|\mu,\sigma^2; a<y<b) = \sigma^2\left[1 + \frac{\left(\frac{a-\mu}{\sigma}\right)\phi\left(\frac{a-\mu}{\sigma}\right) - \left(\frac{b-\mu}{\sigma}\right)\phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)} - \left(\frac{\phi\left(\frac{a-\mu}{\sigma}\right) - \phi\left(\frac{b-\mu}{\sigma}\right)}{\Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)}\right)^2\right] \quad (4)$$

Mean & Variance

For the most common case of a truncation from below at 0, these expression simplify considerably:

$$E(y|\mu,\sigma^2;0<y<\infty)=\mu+\sigma\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}$$

$$\begin{aligned}V(y|\mu,\sigma^2;0<y<\infty)&=\sigma^2\left[1+\frac{\left(\frac{-\mu}{\sigma}\right)\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}-\left(\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}\right)^2\right]\\&=\sigma^2\left[1+\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}\left[\left(-\frac{\mu}{\sigma}\right)-\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}\right]\right]\\&=\sigma^2\left[1-\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}\left[\frac{\phi\left(\frac{\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)}+\left(\frac{\mu}{\sigma}\right)\right]\right]\\&=\sigma^2\left[1-\delta\left(\frac{\mu}{\sigma}\right)\right]\end{aligned}\tag{5}$$

Mean & Variance

In the derivation of the variance expression, we exploit the following properties of the normal distribution: $\phi(\frac{\mu}{\sigma}) = \phi(-\frac{\mu}{\sigma})$ and $1–\Phi(-\frac{\mu}{\sigma}) = \Phi(\frac{\mu}{\sigma})$.

The term $\frac{\phi(\frac{\mu}{\sigma})}{\Phi(\frac{\mu}{\sigma})}$ is often referred to as the "Inverse Mills Ratio" (IMR). The IMR satisfies the following two properties: (1) it is always positive, (2) its derivative w.r.t. (μ/σ) is always negative.

This implies that the truncated expectation (with lower truncation at zero) exceeds the untruncated expectation, but that this effect diminishes the further the untruncated mean is located from the truncation point of zero.

Mean & Variance

Intuitively, the above makes a lot of sense–imagine the normal density with just a tiny bit of its left tail "chopped off" (so it's almost entirely located above zero). This won't affect the (untruncated) expectation very much.

From equation (5) it is also evident that the truncated variance is always smaller than its untruncated counterpart with truncation from below at zero. In fact, this result holds from *any* truncation, lower, upper, or both.

Truncated Regression Model

The truncated regression model (TRM), assuming truncation from below at 0, can be conveniently expressed in terms of latent variable as follows:

$$
\begin{aligned}
z_i &= x_i'\beta + \varepsilon_i, & \varepsilon_i &\sim N(0, \sigma^2) \\
y_i &= \begin{cases} z_i & \text{if } z_i > 0, \\ \text{unobserved} & \text{otherwise.} \end{cases}
\end{aligned}
\tag{6}
$$

The corresponding likelihood function is given by,

$$
L(\beta) = \sigma^{-1} \prod_{i=1}^{n} \left[\Phi\left(\frac{x_i'\beta}{\sigma} \right) \right]^{-1} \phi\left(\frac{y_i - x_i'\beta}{\sigma} \right) \tag{7}
$$

Truncated vs Censored Regression

The key assumption that distinguishes the TRM from the censored regression model (CRM) is that x_i is *unobserved* for $z_i < 0$ in TRM.

This implies that we can't utilize the information that for some individuals on our sample we know that $z_i < 0$–simply because we don't have any such individuals in our sample (most common), or we have them in the sample but we don't have any observable data, dependent or independent, for these individuals (less common).

So how do we know that y_i is truncated if we don't have any individuals in the sample for which we observe x_i, but not y_i? This usually becomes clear from the definition of the dependent variable.

Truncated vs Censored Regression

For e.g., monthly income (abstracting from "negative income" due to credit) must be > 0. Similarly, wages, expenditures, test scores, and other such constructs are usually considered strictly positive.

In many such cases, we ignore the "natural" truncation and forge ahead with a CLRM (OLS etc.), because: (i) we are pretty sure that the entire population distribution lies well above the truncated threshold, or (ii) we transform the dependent variable to be strictly positive, e.g. by taking logs. Thus, a "plain" truncated regression model is fairly uncommon.

Truncated Regression

We first aim to derive estimates for latent model parameters, in this case β. Depending on the context, we may focus on the untruncated or truncated moments for post-estimation.

The *marginal effects* for the truncated model are given as,

$$\frac{\partial E(y_i|z_i > 0)}{\partial x_i} = \beta\left[1 - \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}\left[\left(\frac{x_i'\beta}{\sigma}\right) + \frac{\phi\left(\frac{x_i'\beta}{\sigma}\right)}{\Phi\left(\frac{x_i'\beta}{\sigma}\right)}\right]\right] = \beta\left[1 - \delta\left(\frac{x_i'\beta}{\sigma}\right)\right].$$

Since $\delta\left(\frac{x_i'\beta}{\sigma}\right)$ lies between 0 and 1, this implies the sign of the estimated β is unambiguous, but that the marginal effect is smaller than the estimated coefficients.

Censored Regression Model

The distribution of a censored random variable has two components: (i) a *limit probability*, which maps a segment of the support of the uncensored variable to a single value, and (ii) a *nonlimit density*, which follows the original density for the remainder of the support of the original variable. If the original density is normal with parameterized mean, the Tobit (Type I) model results.

The Tobit model is a composition of the Probit and the standard regression model. In contrast to the TRM, we now observe the explanatory variables for all observations, but the dependent variable only for a subset of the data. This is often referred to as "Censoring".

Censored Regression Model

The larger the unobserved portion, the more important it is to recognize the censoring in your data, as the basic normal regression model would generate misleading results. In most applications, the censoring threshold is again "0". By convention, the missing dependent variables are often coded as zeros in the data set, even though, strictly speaking they are unobserved.

Censored Regression Model

The censored regression model is conveniently expressed as,

$$z_i = x_i'\beta + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$
$$y_i = \begin{cases} z_i & \text{if } z_i > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

Again, the threshold of zero is (an often arbitrary) standard choice and could be changed in a given application. Note that in contrast to the Probit model the error variance is now identified, although poorly if the degree of censoring is high.

Censored Regression Model

The probability of a "0" outcome is explicitly modeled in the likelihood function. It is computed as,

$$\begin{aligned}\Pr(y_i = 0) = \Pr(z_i \leq 0) &= \Pr\left(\frac{\varepsilon_i}{\sigma} \leq \frac{-x_i'\beta}{\sigma}\right) \\ &= \int_{-\infty}^{-x_i'\beta/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2\right] d\varepsilon_i \\ &= \Phi\left(-\frac{x_i'\beta}{\sigma}\right)\end{aligned} \tag{9}$$

For observed values or the nonlimit density, we have the normal distribution $y_i \sim N(x_i'\beta, \sigma^2)$.

Unconditional Expectation

The unconditional expectation for a censored normal dependent variable, assuming censoring from below at some constant "a", can be computed as follows.

$$\begin{aligned} E(y_i) &= \Pr(y_i = a)E(y_i|y_i = a) + \Pr(y_i > a)E(y_i|y_i > a) \\ &= \Pr(z_i \leq a)a + \Pr(z_i > a)E(z_i|z_i > a) \\ &= \Phi\Big(\frac{a - x_i'\beta}{\sigma}\Big) * a + \Phi\Big(\frac{-a + x_i'\beta}{\sigma}\Big)\left[x_i'\beta + \sigma\frac{\phi\big(\frac{-a+x_i'\beta}{\sigma}\big)}{\Phi\big(\frac{-a+x_i'\beta}{\sigma}\big)}\right] \end{aligned} \tag{10}$$

It is evident from our discussion of the truncated regression model, the last component in equation (17) is the conditional or truncated expectation, i.e., $E(y_i|y_i > 0)$. Either expectation may be relevant in a given application.

Likelihood

The likelihood function can be expressed as,

$$
\begin{aligned}
L(\beta, \sigma^2) &= \prod_{i:y_i=0} \Phi\Big(-\frac{x_i'\beta}{\sigma}\Big) \prod_{i:y_i>0} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(y_i - x_i'\beta)^2\right] \\
&= \prod_{i:y_i=0} \Phi\Big(-\frac{x_i'\beta}{\sigma}\Big) \prod_{i:y_i>0} \frac{1}{\sigma}\phi\Big(\frac{y_i - x_i'\beta}{\sigma}\Big) \qquad (11) \\
&= \prod_{i=1}^{n} \left[\Phi\Big(-\frac{x_i'\beta}{\sigma}\Big) I(y_i = 0) + \frac{1}{\sigma}\phi\Big(\frac{y_i - x_i'\beta}{\sigma}\Big) I(y_i > 0)\right]
\end{aligned}
$$

Posterior Quantities

The two primary posterior constructs of interest in the Tobit model are usually the expected value of the outcome variable under censoring (i.e. the conditional or truncated expectation mentioned above), and the marginal effect of a regressor on this expectation.

$$
\begin{aligned}
E(y_p|x_p, z_p > 0) &= x_p'\beta + \sigma\frac{\phi\left(\frac{x_p'\beta}{\sigma}\right)}{\Phi\left(\frac{x_p'\beta}{\sigma}\right)} \\
\frac{\partial E(y_p|x_p)}{\partial x_j}\bigg| z_p > 0 &= \beta_j\Phi\left(\frac{\bar{x}_p'\beta}{\sigma}\right)
\end{aligned}
\tag{12}
$$

where x_p is usually set to the sample mean $\bar{x}_p$ for the derivation of the marginal effects.

Posterior Quantities

If x_j is an indicator variable, it's marginal effect is expressed more meaningfully as,

$$\begin{aligned} & E(y_p|z_p > 0, \bar{x}_{-j}, x_j = 1) - E(y_p|z_p > 0, \bar{x}_{-j}, x_j = 0) \\ &= \left[\bar{x}_p'\beta + \sigma\frac{\phi\left(\frac{\bar{x}_p'\beta}{\sigma}\right)}{\Phi\left(\frac{\bar{x}_p'\beta}{\sigma}\right)}\middle| x_j = 1\right] - \left[\bar{x}_p'\beta + \sigma\frac{\phi\left(\frac{\bar{x}_p'\beta}{\sigma}\right)}{\Phi\left(\frac{\bar{x}_p'\beta}{\sigma}\right)}\middle| x_j = 0\right]. \end{aligned} \tag{13}$$

Posterior Quantities

For the Tobit model, it is also meaningful to examine the marginal effect of regressors on the latent variable z. For example, consider an auction for adopting/buying an horse. With an imposed minimum bid, it is feasible that a potential buyer's willingness to pay (WTP) lies below the threshold. In other words, the latent variable is well defined, even in the unobserved realm.

Thus, one may want to examine a regressor's effect on "latent WTP", regardless of passing the threshold or not. In that case, we can interpret the regression coefficients as in the standard regression model.

Ignoring Censoring

One might be tempted to simply ignore the censoring in the dependent variable and treat the threshold observations (usually "zeros") in the same way as all other outcome values. However, the resulting simple regression estimates are biased, with the bias increasing with the degree of censoring.

Running a basic regression implies that you are setting the censoring threshold to $-\infty$ (i.e., no censoring). This would lead to the usual form of the unconditional expectation:

$$E(y_p|x_p, z_p > -\infty) = x_p'\beta + \sigma\frac{\phi(\frac{\infty + x_p'\beta}{\sigma})}{\Phi(\frac{\infty + x_p'\beta}{\sigma})} = x_p'\beta + \frac{0}{1} = x_p'\beta. \quad (14)$$

Ignoring Censoring

However, if the censoring threshold is not $-\infty$, but say some value a, the unconditional expectation will be biased downwards compared to the true conditional expectation since the Inverse Mills Ratio $\phi(\frac{a+x_p'\beta}{\sigma})/\Phi(\frac{a+x_p'\beta}{\sigma})$ is always positive.

Moreover, as evident from the (correct) marginal effect in equation (12), the coefficient estimates will be biased towards zero, or–equivalently put–biased downward in absolute terms for the misspecified model.

Ignoring Censoring

The good news is that the Inverse Mills Ratio decreases in $x_p'\beta$. Thus, as the unconditional expectation $x_p'\beta$ increases, the bias diminishes. Intuitively, this means that if the unconditional mean is far to the right from the censoring threshold, the censoring effect is small and ignoring it is less damaging.

Thank you!