Econ 526 | Prof. Carlos A. Flores

Winter 2016 | Cal Poly State University

NOTES V: Truncated and Censored Dependent Variables

(Ruud Chap. 28, Greene Chap. 20)

Version: February 12$^{\text{th}}$, 2016

Consider the classical linear regression model where $y_i = x_i'\beta_0 + \sigma_0\varepsilon_i$ and $\varepsilon_i \backsim N(0,1)$. In these notes we analyze cases in which the dependent variable is constrained in some way. Specifically, we say that the dependent variable is *truncated* if relevant observations outside a specified range are completely lost; and we say that the dependent variable is *censored* if its value is cut off outside a certain range with several observations at the endpoints of this range. For example, suppose we want to analyze the effect of a set of covariates on earnings. If our sample consists of only people earning below \$100,000 a year, then our dependent variable is truncated at that value. In other words, we do not have a random sample of the population but, at most, a random sample conditional on earnings being less than \$100,000. On the other hand, suppose that our data on earnings came from the Social Security Administration (SSA) earning records. As explained in Chay and Powell (2001), "anyone earning more than the maximum that was taxable under Social Security is recorded as having earned at the maximum". Hence, in this case our dependent variable is censored at the taxable maximum. Suppose this taxable maximum were \$100,000, then, our observed y_i would reflect actual earnings if earning are less than \$100,000, but it would be equal to \$100,000 for all those earning above \$100,000.

Note that when the dependent variable is censored we have more information than when is truncated. When y_i is censored we know something about its value (e.g., is above \$100,000) and we also observe its covariates. On the other hand, when y_i is truncated we do not observe anything, so all its information is missing.

The problem of having a limited dependent variable is that regressing y_i on x_i will lead to bias in the estimates of β_0. To gain intuition into why this is the case consider estimating a linear regression of income (y_i) on education (x_i), and suppose that income is censored at \$100,000 a year. Figure 1 plots some representative data for this case. The observations shown with an "x" represent the data we would use if the earning variable were not censored, and the observations shown with a "o" represent our censored (and available) data. Figure 1 depicts the estimated regression lines when using each of those data. Clearly, the model we estimate using our censored data is biased and underestimates the true slope coefficient.

Our goal in these notes is to present models that allow us to obtain consistent estimates of β_0 in cases where our dependent variable is truncated or censored. The models presented here are based on the assumption that the error term ε_i is normally distributed. Assuming that the error density is known will allow us to use maximum likelihood estimation and obtain asymptotically efficient estimators of β_0. On the other hand, as usual with ML models, if our assumption on the distribution of the error term is incorrect, our estimators of β_0 will be inconsistent. This has motivated the study of semiparametric estimators of truncated and censored models, which do not require specific assumptions about the distribution of the error term[1]. However, estimation of these models under the normality assumption remains a very used approach in applied econometrics and also provides a good introduction to the topic.

1 Useful facts about the normal distribution.

Before we start discussing truncated and censored models into more detail is important to consider a few useful results to be used later.

Result 1. If $x \backsim f(x)$, then the density of x given $x > a$ is $f(x|x > a) = f(x) / \Pr(x > a)$. Similarly, $f(x|x < a) = f(x) / \Pr(x < a)$.

[1]See, for example, the survey in Chay and Powell (2001).

This result follows directly from the definition of a conditional probability. The idea is that we have to make the density $f(x|x>a)$ integrate to 1, and the way to do it is by scaling $f(x)$ up by $\Pr(x>a)$ so that:

$$\int_{x>a} f(x|x>a)\,dx = \int_{x>a} \frac{f(x)}{\Pr(x>a)}dx = \frac{1}{\Pr(x>a)} \int_{x>a} f(x)\,dx = 1$$

Figure 2 shows a graph of the conditional density x given $x > a$. Note how $f(x)$ is re-scaled so that it integrates to 1.

Given that we will assume that the error term in the regression function follows a normal distribution, it is important to consider some useful facts about this distribution.

Result 2. Let $z \backsim N(0,1)$ and let $\phi(z)$ and $\Phi(k) = \Pr(z \leq k)$ be the density and cumulative functions, respectively, of the standard normal distribution. Then,

a. $\Pr(z \geq k) = 1 - \Phi(k) = \Phi(-k)$ and $\phi(z) = \phi(-z)$

b. $\frac{\partial \phi(z)}{\partial z} = -z\phi(z)$

c. $E[z|z<k] = -\frac{\phi(k)}{\Phi(k)}$

d. $E[z|z>k] = \frac{\phi(k)}{1-\Phi(k)} = \frac{\phi(-k)}{\Phi(-k)}$

The first result (2.a) follows directly from the symmetry of the normal distribution. The second one is straightforward to show by deriving the density of the standard normal distribution. The third result can be shown as follows

$$\begin{aligned} E[z|z<k] &= \int_{-\infty}^{k} zf(z|z>k)\,dz = \int_{-\infty}^{k} z\frac{f(z)}{\Pr(z<k)}dz \\ &= \frac{1}{\Phi(k)} \int_{-\infty}^{k} z\phi(z)\,dz = -\left.\frac{\phi(z)}{\Phi(k)}\right|_{-\infty}^{k} = -\frac{\phi(k)}{\Phi(k)} \end{aligned} \tag{1}$$

where in the first line we used the definitions of expectation and the conditional distribution $f(z|z>k)$ (see result 1); and in the third line we used result 2.b. Result 2.d can be shown using the same steps as in (1).

Result 3. Let $u \backsim N(0, \sigma^2)$. Then,

a. $E[u|u<k] = -\sigma\frac{\phi(k/\sigma)}{\Phi(k/\sigma)}$

b. $E[u|u>k] = \sigma\frac{\phi(k/\sigma)}{1-\Phi(k/\sigma)} = \sigma\frac{\phi(-k/\sigma)}{\Phi(-k/\sigma)}$

These results follow directly from results 2.c and 2.d. For example, note that $u = \sigma z$ so we can write

$$E[u|u<k] = E[\sigma z|\sigma z<k] = \sigma E[z|z<k/\sigma] = -\sigma\frac{\phi(k/\sigma)}{\Phi(k/\sigma)}$$

where in the last equality we used result 2.c.

2 Truncated Models.

In this section we consider the case in which some observations are totally missing when the value of the dependent variable falls in some range. Let y_i^* be a latent variable such that

$$y_i^* = x_i'\beta_0 + \sigma_0\varepsilon_i \tag{2}$$

where $\varepsilon_i \backsim N(0,1)$, x_i is a $k \times 1$ vector, and x_i and ε_i are independent. Our interest is on $E[y_i^*|x] = x_i'\beta_0$, and in particular on the parameter β_0. However, instead of observing a sample of (y_i^*, x_i), we observe a sample size n of (y_i, x_i), where y_i can take the form

$$y_i = \begin{cases} \text{unobserved if } y_i^* < a \\ y_i^* \text{ if } a \leq y_i^* \leq b \\ \text{unobserved if } b < y_i^* \end{cases} \tag{3}$$

with $a < b$. The case in which both a and b are finite numbers is called "interval truncation". When $a = -\infty$ ($a > -\infty$) and $b < \infty$ ($b = \infty$) our variable is truncated from the right (left). Figure 3 illustrates each of these cases. The example discussed in the introduction, in which our dependent variable is earnings and is truncated at \$100,000, is an example of right truncation.

In this section, we focus on the case of left truncation and assume the truncation occurs at 0. The model in (3) and more complicated models of truncation can be handled in the same way as the case we consider here. Assume we observe a sample size n of (y_i, x_i), where y_i is given by

$$y_i = \begin{cases} y_i^* \text{ if } y_i^* \geq 0 \\ \text{unobserved if } y_i^* < 0 \end{cases} \tag{4}$$

Suppose we ignored the fact that our dependent variable is truncated and regressed y_i on x_i. Then, the conditional expectation of y_i given x_i *and* the fact that we observed y_i is given by:

$$\begin{aligned} E\left[y_i | x_i, y_i \text{ is observed}\right] &= E\left[y_i^* | x_i, y_i^* \geq 0\right] \\ &= E\left[x_i'\beta_0 + \sigma_0\varepsilon_i | x_i, y_i^* \geq 0\right] \\ &= x_i'\beta_0 + \sigma_0 E\left[\varepsilon_i | y_i^* \geq 0\right] \\ &= x_i'\beta_0 + \sigma_0 E\left[\varepsilon_i | x_i'\beta_0 + \sigma_0\varepsilon_i \geq 0\right] \\ &= x_i'\beta_0 + \sigma_0 E\left[\varepsilon_i | \varepsilon_i \geq -\frac{x_i'\beta_0}{\sigma_0}\right] \\ &= x_i'\beta_0 + \sigma_0 \frac{\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)}{\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)} \\ &= x_i'\beta_0 + \sigma_0\lambda\left(\frac{x_i'\beta_0}{\sigma_0}\right) \end{aligned} \tag{5}$$

where in the first two lines we used the definitions of y_i and y_i^*, respectively. In the third line we used the fact that x_i and ε_i are assumed independent, and in the second to last line we used result 2.d. Finally, we defined $\lambda\left(\frac{x_i'\beta_0}{\sigma_0}\right) = \phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) / \Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)$. This term is usually referred as the "inverse Mills' ratio".

From (5) we can see that if we regress y_i on x_i we would overestimate the conditional expectation (i.e., the true regression line) because the second term to right of (5) is always positive. In addition, it can be shown (e.g., Greene chapter 20 or Ruud chapter 28) that the estimated coefficients (excluding the intercept) from an OLS of y_i on x_i using the observed data will be biased *towards zero*. Note that we can look at the problem of ignoring the truncation in our dependent variable as an omitted variable bias, where the omitted variable in this case is the inverse Mills' ratio.[2] Finally, note that the inverse Mills' ratio is a nonlinear function of the covariates.

Given that we are assuming that the distribution of the error terms is known, we use a ML approach, which we know yields asymptotically efficient estimators under our assumptions. As in any ML problem, the critical step is writing the log-likelihood function. Remember that the log-likelihood $L(\theta|y_i, x_i)$ is given by

$$L(\theta|y_i, x_i) = \sum_{i=1}^{n} \log\left(f_y\left(y_i|\theta, x_i\right)\right) \tag{6}$$

In our case we have that $\theta = \{\beta, \sigma\}$. Now, we focus on finding $f_y(y_i|\theta, x_i)$. The key to obtain the density of y_i is to note that, by its definition in (4), y_i is equal to y_i^* conditional on $y_i^* \geq 0$. Hence, we have that $f_y(y_i|\theta, x_i) = f_{y^*}(y^*|y_i^* \geq 0, \theta, x_i)$. Moreover, by using result 1 we have that

$$f_{y^*}\left(y^*|y_i^* \geq 0, \theta, x_i\right) = \frac{f_{y^*}\left(y^*|\theta, x_i\right)}{\Pr\left(y_i^* \geq 0|\theta, x_i\right)} \tag{7}$$

We now need expressions for the terms in (7). We start by deriving an expression for $f_{y^*}(y^*|\theta, x_i)$. To simplify notation, in what follows we drop the conditional arguments θ, x_i. Using our assumption that ε_i follows a standard normal distribution we can write

$$\begin{aligned} F\left(y^*\right) &= \Pr\left(y_i^* \leq y^*\right) = \Pr\left(x_i'\beta_0 + \sigma_0\varepsilon_i \leq y^*\right) \\ &= \Pr\left(\varepsilon_i \leq \frac{y^* - x_i'\beta_0}{\sigma_0}\right) = \Phi\left(\frac{y^* - x_i'\beta_0}{\sigma_0}\right) \end{aligned} \tag{8}$$

Therefore,

[2] Also, one can show that the variance of the OLS model in this case is heteroskedastic, that is, $Var\left[y_i|x_i, y_i \text{ is observed}\right]$ is not constant. Therefore, the estimated standard errors from this regression would also be inconsistent.

$$f_{y^*}(y^*|\theta, x_i) = \frac{\partial F(y^*)}{\partial y^*} = \frac{1}{\sigma_0}\phi\left(\frac{y^* - x_i'\beta_0}{\sigma_0}\right) \tag{9}$$

Now we find an expression for $\Pr(y_i^* \geq 0|\theta, x_i)$. Following similar steps as in (8) we find

$$\Pr(y_i^* \geq 0|\theta, x_i) = 1 - \Pr(y_i^* < 0|\theta, x_i) = 1 - \Phi\left(-\frac{x_i'\beta_0}{\sigma_0}\right) \tag{10}$$

Plugging equations (9) and (10) into (7), and using the fact that $y_i = y_i^*$ conditional on $y_i^* \geq 0$ we find

$$f_y(y_i|\theta, x_i) = f_{y^*}(y^*|y_i^* \geq 0, \theta, x_i) = \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)}{1 - \Phi\left(-\frac{x_i'\beta}{\sigma}\right)} \tag{11}$$

Finally, if we plug (11) into (6) we find that the log-likelihood for the truncated regression model in (4) is given by

$$L(\beta, \sigma|y_i, x_i) = \sum_{i=1}^{n}\left\{\log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] - \log\left[1 - \Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right]\right\} \tag{12}$$

This likelihood is maximized with respect to β and σ to get estimates $\widehat{\beta}$ and $\widehat{\sigma}$. Similar to other models we have seen before (e.g., probit), there is no closed form solution for our estimators $\widehat{\beta}$ and $\widehat{\sigma}$, so numerical methods as those previously discussed must be used (e.g., Newton-Raphson or BHHH). Unfortunately, the likelihood in (12) is not always well behaved and may be hard to maximize. A useful reparametrization proposed by Olsen (1978) is to let $\gamma = \beta/\sigma$ and $\delta = 1/\sigma$ so that the likelihood is given by

$$L(\gamma, \delta|y_i, x_i) = \sum_{i=1}^{n}\left\{\log\left[\delta\phi\left(\delta y_i - x_i'\gamma\right)\right] - \log\left[1 - \Phi\left(x_i'\gamma\right)\right]\right\} \tag{13}$$

and maximized with respect to γ and δ. This likelihood function tends to be better behaved than the one in (12); in other words, it is easier to find $\widehat{\gamma}$ and $\widehat{\delta}$ than $\widehat{\beta}$ and $\widehat{\sigma}$. We can recover $\widehat{\beta}$ and $\widehat{\sigma}$ from $\widehat{\sigma} = 1/\widehat{\delta}$ and $\widehat{\beta} = \widehat{\delta}/\widehat{\gamma}$, and derive their asymptotic variance by using the delta method. As usual, under the assumptions of this model, $\widehat{\beta}$ and $\widehat{\sigma}$ are consistent, asymptotically normal and efficient (they are MLE estimators).

3 Censored Models.

In this section we analyze the case in which the dependent variable is censored, so that its value is cut off outside a certain range. The general set up is basically the same we used for truncation models. As in the previous section, let y_i^* be a latent variable such that

$$y_i^* = x_i'\beta_0 + \sigma_0\varepsilon_i \tag{14}$$

where $\varepsilon_i \backsim N(0,1)$, x_i is a $k \times 1$ vector, and x_i and ε_i are independent. As before, our interest is on estimation of the parameter β_0. However, instead of observing a sample of (y_i^*, x_i), we observe a sample size n of (y_i, x_i), where y_i can now take the form

$$y_i = \begin{cases} a \text{ if } y_i^* \leq a \\ y_i^* \text{ if } a < y_i^* < b \\ b \text{ if } b \leq y_i^* \end{cases} \tag{15}$$

with $a < b$. The case in which both a and b are finite numbers is called "interval censoring". When $a = -\infty$ $(a > -\infty)$ and $b < \infty$ $(b = \infty)$ our variable is censored from the right (left). Figure 4 illustrates each of these cases. The example discussed in the introduction of our notes, in which our dependent variable is earnings and is equal to \$100,000 if earnings are equal to or greater than \$100,000, is an example of right censoring. The case when $a = 0$ and $b = \infty$ was first proposed by Tobin (1958) when analyzing consumers' purchases of durable goods. Because of this, econometricians usually refer to censored regression models as "Tobit" models.[3]

In this section we focus on the original Tobit model, which corresponds to the case of

[3] As you will see, econometricians have a tendency to "it" models like this, mainly because of their similarity to probit models. Then, you have that the model proposed by Tobin is usually called a Tobit model, and the sample selection model proposed by Heckman (which we will see later) is usually called "Heckit".

left censoring at zero. More general types of censoring can be approached in the same way. Assume that we have a sample size n of (y_i, x_i), where y_i is given by

$$y_i = \begin{cases} 0 \text{ if } y_i^* \leq 0 \\ y_i^* \text{ if } y_i^* > 0 \end{cases} \tag{16}$$

or $y_i = \max\{0, y_i^*\}$. As in the previous section, we start by supposing we ignored the fact that our dependent variable is censored and regressed y_i on x_i. Then, it can be shown (e.g., Greene Chap. 20 or Ruud Chap 28) that the conditional expectation of y_i given x_i is given by:

$$E[y_i|x_i] = x_i'\beta_0\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) + \sigma_0\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) \tag{17}$$

It is clear from (17) that our OLS estimate of β_0 will be biased and inconsistent because of $\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)$ being different from one and because of the omitted variable $\phi\left(\frac{x_i'\beta_0}{\sigma_0}\right)$. In fact, it will typically be the case that the coefficient estimates will be biased toward zero.[4,5]

As with truncated models, we take advantage of our normality assumption on the error term and use a ML approach to derive consistent and asymptotically efficient estimators of β_0. We now proceed to derive the log-likelihood function for the Tobit model in (16). The objective is to find the density (or probability function) $f_y(y_i|\beta, \sigma, x_i)$ [6] so that we can write the log-likelihood function using (6). By (16), note that y_i can be in one of two groups: can be either equal to y_i^* if $y_i^* > 0$, or equal to 0 if $y_i^* \leq 0$. Then, the likelihood (or loosely speaking, the probability) of observing a particular value of y_i is equal to the probability of observing y_i^* if $y_i^* > 0$; and the likelihood of observing $y_i = 0$ is equal to the probability that $y_i^* \leq 0$. Hence, we have that

[4] Intuitively, $\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) < 1$ so $\beta_0\Phi\left(\frac{x_i'\beta_0}{\sigma_0}\right) < \beta_0$.

[5] As in the truncated model, one can show that the variance of the OLS model in this case is heteroskedastic, so that the estimated standard errors from this regression would also be inconsistent.

[6] Or equivalently, find the likelihood function for observation i given by $f_y(\beta, \sigma, y_i|x_i)$.

$$f_y\left(y_i|\beta,\sigma,x_i\right)=\begin{cases}\Pr\left(y_i^*\le 0\right) & \text{if } y_i=0\\ f_{y^*}\left(y_i^*|\beta,\sigma,x_i\right) & \text{if } y_i>0\end{cases} \tag{18}$$

Note that $f_y\left(y_i|\beta,\sigma,x_i\right)$ has a continuous and a discrete part. This type of distribution is usually called a "mixed probability function". Figure 5 shows a graph of the probability function in (18).

We now get expressions for the functions in the right side of (18). Using steps similar to the ones in (8) we find

$$\begin{aligned}\Pr\left(y_i=0\right) &= \Pr\left(y_i^*\le 0\right)=\Pr\left(x_i'\beta_0+\sigma_0\varepsilon_i\le 0\right)\\ &= \Pr\left(\varepsilon_i\le -\frac{x_i'\beta_0}{\sigma_0}\right)=\Phi\left(-\frac{x_i'\beta_0}{\sigma_0}\right)\end{aligned} \tag{19}$$

Also, from (9) we have that $f_{y^*}\left(y^*|\beta,\sigma,x_i\right)=\frac{1}{\sigma_0}\phi\left(\frac{y^*-x_i'\beta_0}{\sigma_0}\right)$. Therefore, we can write the likelihood for observation i as

$$f_y\left(\beta,\sigma,y_i|x_i\right)=\begin{cases}\Phi\left(-\frac{x_i'\beta}{\sigma}\right) & \text{if } y_i=0\\ \frac{1}{\sigma}\phi\left(\frac{y_i-x_i'\beta}{\sigma}\right) & \text{if } y_i>0\end{cases} \tag{20}$$

Finally, let $d_i=1$ if y_i is not censored (i.e., if $y_i^*>0$) and let $d_i=0$ if y_i is censored (i.e., if $y_i^*\le 0$). Then, we can write (20) as

$$f_y\left(\beta,\sigma,y_i|x_i\right)=\left[\frac{1}{\sigma}\phi\left(\frac{y_i-x_i'\beta}{\sigma}\right)\right]^{d_i}\left[\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right]^{1-d_i} \tag{21}$$

and the log-likelihood function for the censored regression model in (16) as

$$L\left(\beta,\sigma|y_i,x_i\right)=\sum_{i=1}^{n}\left\{d_i\log\left[\frac{1}{\sigma}\phi\left(\frac{y_i-x_i'\beta}{\sigma}\right)\right]+\left(1-d_i\right)\log\left[\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right]\right\} \tag{22}$$

In the Tobit model the log-likelihood is maximized with respect to β and σ to obtain consistent and asymptotically efficient estimators $\widehat{\beta}$ and $\widehat{\sigma}$. As in the truncated case, there are no closed form expressions for $\widehat{\beta}$ and $\widehat{\sigma}$, so numerical methods are needed for their computation. Similarly, Olsen's reparametrization ($\gamma=\beta/\sigma$, $\delta=1/\sigma$) simplifies computation a lot. In fact, it can be shown (e.g., Lemma 28.3 in Ruud p. 813) that under the normality assumption the reparametrized version of the censored log-likelihood function is globally concave. Therefore, the numerical methods previously discussed (e.g., Newton-Raphson and BHHH) work very well in this case.

4 Comparison of OLS, Probit, and truncated and censored models.

To gain more intuition on truncated and censored models, in this section we compare their general set up and log-likelihoods to the ones from OLS and probit models. As before, let y_i^* be a latent variable such that

$$y_i^* = x_i'\beta_0 + \sigma_0\varepsilon_i \tag{23}$$

where $\varepsilon_i \backsim N(0,1)$, x_i is a $k \times 1$ vector, and x_i and ε_i are independent. For each of the four models previously mentioned, the type of data we observe can be written as:

$$\begin{aligned}
\text{OLS} &: \quad y_i = y_i^* \\
\text{Probit} &: \quad y_i = \begin{cases} 0 \text{ if } y_i^* \leq 0 \\ 1 \text{ if } y_i^* > 0 \end{cases} \\
\text{Censored} &: \quad y_i = \begin{cases} 0 \text{ if } y_i^* \leq 0 \\ y_i^* \text{ if } y_i^* > 0 \end{cases} \\
\text{Truncated} &: \quad y_i = \begin{cases} \text{unobserved if } y_i^* < 0 \\ y_i^* \text{ if } y_i^* \geq 0 \end{cases}
\end{aligned} \tag{24}$$

Similarly, under the normality assumption, the log-likelihood for each of these models can be written as

$$
\begin{aligned}
\text{OLS} &: \quad L(\beta,\sigma|\cdot) = \sum_{i=1}^{n} \log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] && (25)\\
\text{Probit} &: \quad L(\beta|\cdot) = \sum_{i=1}^{n} \left\{y_i \log\left[\Phi\left(x_i'\beta\right)\right] + (1-y_i)\log\left[\Phi\left(-x_i'\beta\right)\right]\right\}\\
\text{Censored} &: \quad L(\beta,\sigma|\cdot) = \sum_{i=1}^{n} \left\{d_i \log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] + (1-d_i)\log\left[\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right]\right\}\\
\text{Truncated} &: \quad L(\beta,\sigma|\cdot) = \sum_{i=1}^{n} \left\{\log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] - \log\left[1-\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right]\right\}
\end{aligned}
$$

where for the censored model

$$
d_i = \begin{cases} 0 \text{ if } y_i^* \leq 0 \\ 1 \text{ if } y_i^* > 0 \end{cases} \tag{26}
$$

Also, remember that the scale parameter is not identified in the probit model, so we set $\sigma = 1$. Figure (6) shows the probability (density) function for each of these four models. Note that while in the OLS model we fully observe the latent variable y_i^*, in the probit model we only observe whether y_i^* falls in one of two regions. Hence, we do not have information on the scale of y_i^* and σ is not identified.

By comparing figures 6.a and 6.d, along with the log-likelihoods for the OLS and truncated models, we can think of the truncated regression model as a "re-scaled" version of the OLS, in which we re-scale the density of the observed data so that it integrates to one. If we look at the censored log-likelihood we can see that it looks like a combination of an OLS and a probit log-likelihood, where the OLS part is used for the uncensored data and the probit part is used for the censored data. In fact, if we add and subtract $d_i \log\left[\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]$ to the inside part of the log-likelihood of the censored model we find

$$
\begin{aligned}
L(\beta,\sigma|\cdot) &= \sum_{i=1}^{n} \left\{d_i \left(\log\left[\frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right] - \log\left[\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right]\right)\right. && (27)\\
&\quad \left. + d_i \log\left[\Phi\left(\frac{x_i'\beta}{\sigma}\right)\right] + (1-d_i)\log\left[\Phi\left(-\frac{x_i'\beta}{\sigma}\right)\right]\right\} && (28)
\end{aligned}
$$

Therefore, the log–likelihood for the censored model can also been seen as a combination of the truncated log-likelihood (27) for the uncensored observations (i.e., those with $d_i = 1$) and a probit log-likelihood for the indicator variable of whether the observation is censored or uncensored (28).

5 Final notes on Truncated and Censored models.

In these notes, we have discussed consistent and asymptotically efficient estimators for truncated and censored models. Of course, the asymptotic efficiency result is not achieved for free, but by assuming that the density of the error term is known. Specifically, we assumed throughout these notes that $\varepsilon_i \backsim N(0,1)$. The cost of making such a strong assumption is that if the true distribution of the error term is not normal and/or if is heteroskedastic, then our estimator of $\widehat{\beta}$ will be inconsistent. Moreover, even if the distribution of the error term is correctly specified, maximum likelihood estimation of censored models when using panel data with fixed effects will lead to inconsistent estimates of β_0.[7] Because of this, econometricians have looked at semiparametric methods for estimation of β_0 when the data is censored and/or truncated. These models do not require specification of the error distribution, and in that sense, are consistent under more general conditions than the ML estimators presented in these notes. A nice and non-technical survey of these methods can be found in Chay and Powell (2001).

References

[1] Chay, K. Y. and Powell, J. (2001). "Semiparametric Censored Regression Models", *Journal of the Economic Perspectives,* 15(4), 29-42.

[7] A panel data is one where we have repeated observations over time. Time permiting, we will study the fixed effects model of panel data later in the course. This model is widely used in applied econometrics when using panel data.

[2] Olsen, R.J. (1978). "Note on the Uniqueness of the Maximum Likelihood Estimator for the Tobit Model", *Econometrica,* 46(5), 1211-1215.

FIGURE 1

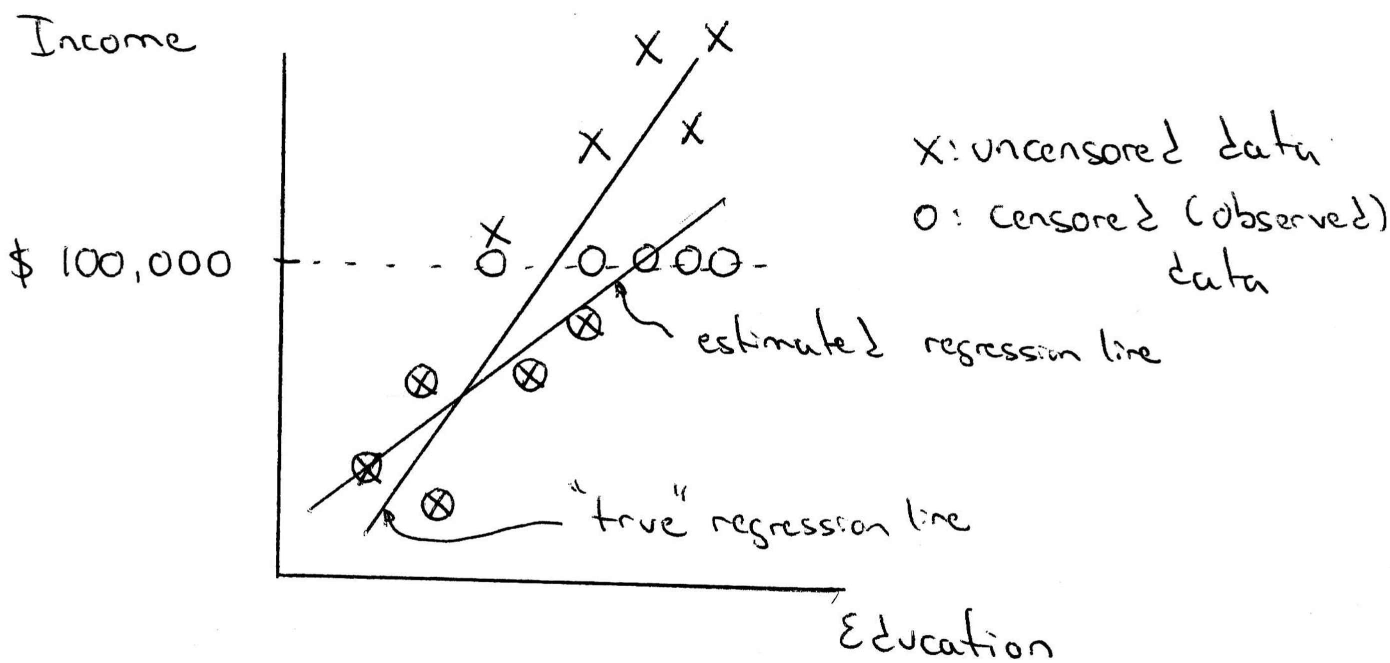


FIGURE 2

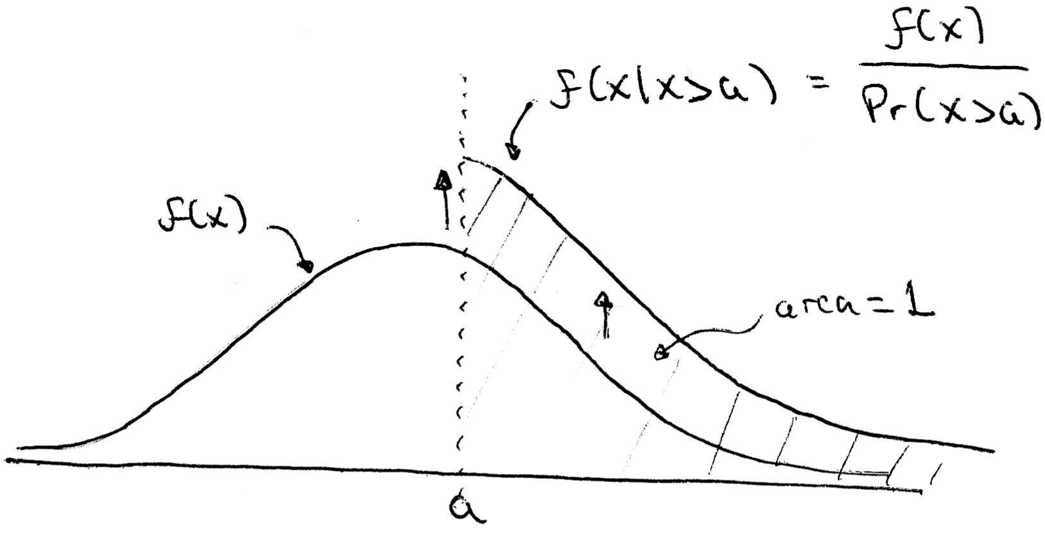
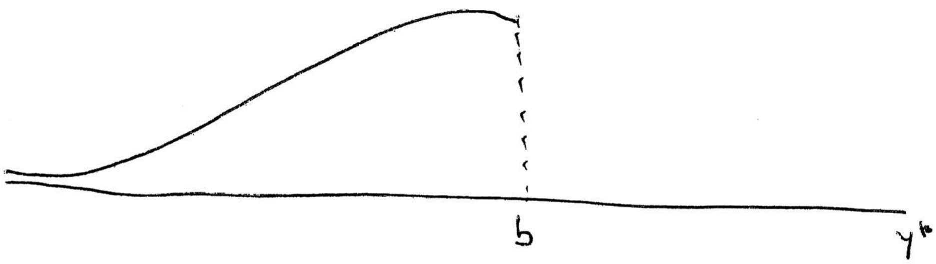
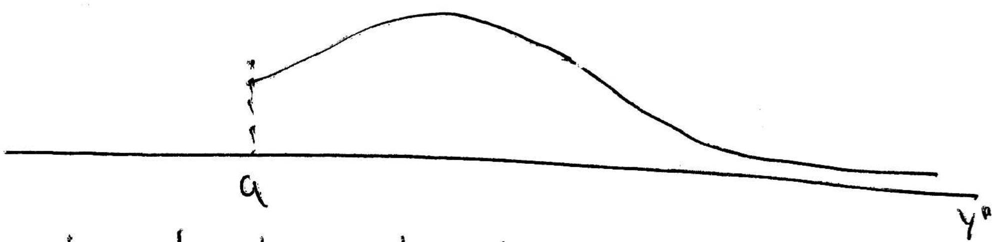


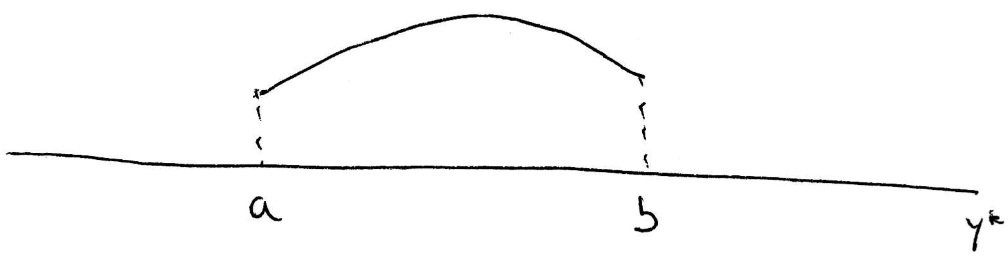
FIGURE 3 & 4



Right truncation/censoring



Left truncation/censoring



Interval truncation/censoring

FIGURE 5

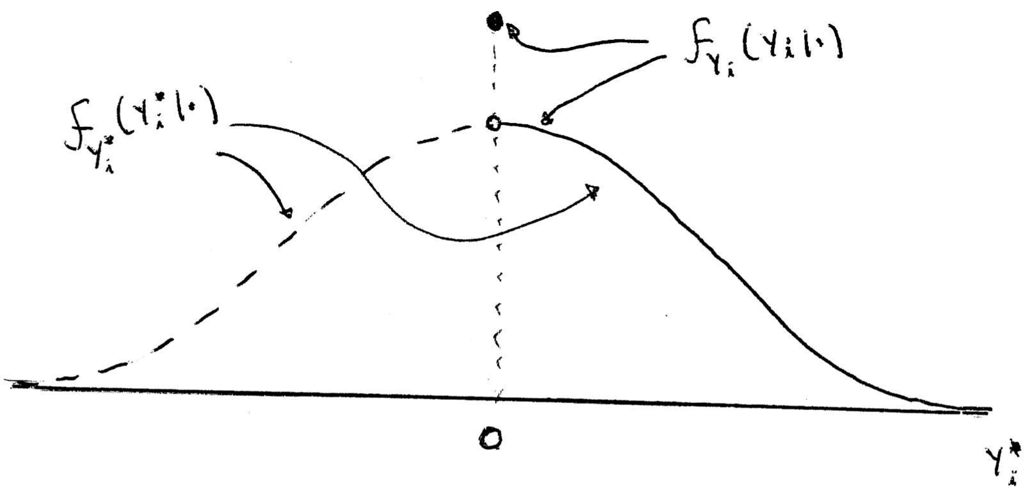
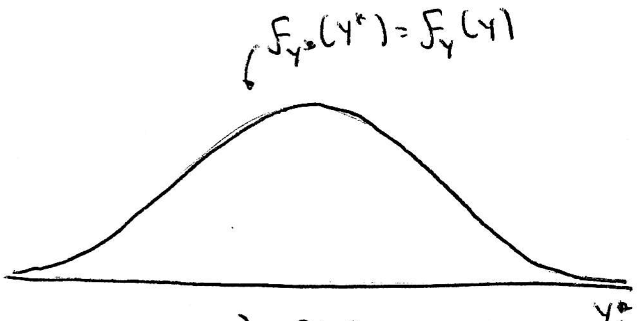
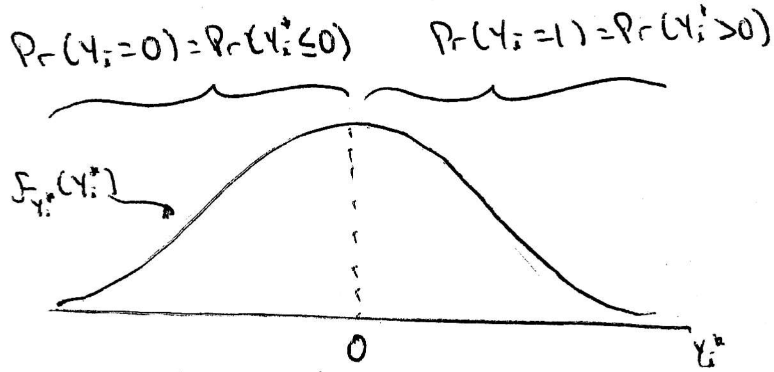


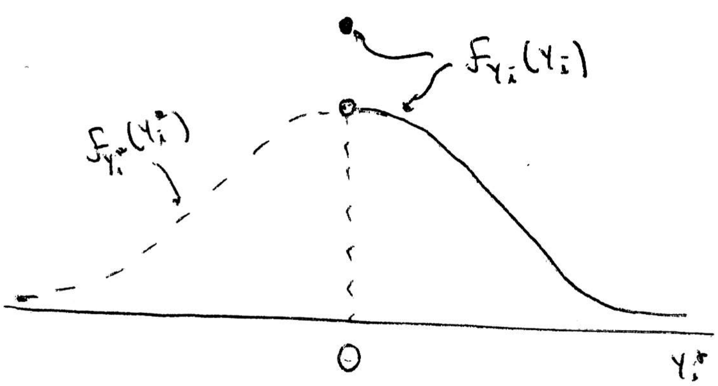
FIGURE 6



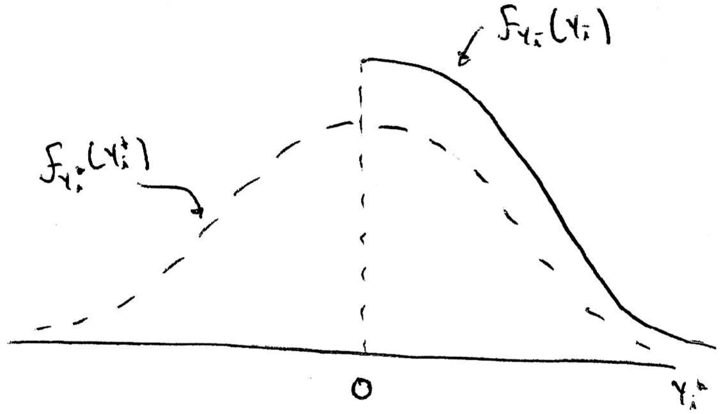
a)- OLS



b)- Probit



c)- Censored model



d)- Truncated model