7 Limited Outcomes: The Tobit Model

In the linear regression model, the values of all variables are known for the entire sample. This chapter considers the situation in which the sample is limited by censoring or truncation. *Censoring* occurs when we observe the independent variables for the entire sample, but for some observations we have only limited information about the dependent variable. For example, we might know that the dependent variable is less than 100, but not know how much less. *Truncation* limits the data more severely by excluding observations based on characteristics of the dependent variable. For example, in a truncated sample all *cases* where the dependent variable is less than 100 would be deleted. While truncation changes the sample, censoring does not.

The classic example of censoring is Tobin's (1958) study of household expenditures. A consumer maximizes utility by purchasing durable goods under the constraint that total expenditures do not exceed income. Expenditures for durable goods must at least equal the cost of the least expensive item. If a consumer has only $50 left after other expenses and the least expensive item costs $100, the consumer can spend nothing on durable goods. The outcome is censored since we do not know how much a household would have spent if a durable good could be purchased for less than $100. Many other examples of censored outcomes

can be found: hours worked by wives (Quester & Greene, 1982), scientific publications (Stephan & Levin, 1992), extramarital affairs (Fair, 1978), foreign trade and investment (Eaton & Tamura, 1994), austerity protests in Third World countries (Walton & Ragin, 1990), damage caused by a hurricane (Fronstin & Holtmann, 1994), and IRA contributions (LeClere, 1994). Amemiya (1985, p. 365) lists many additional examples.

Hausman and Wise's (1977) analysis of the New Jersey Negative Income Tax Experiment is an early application of models for truncated data. In this study, families with incomes more than 1.5 times the poverty level were excluded from the sample. Thus, the sample itself is affected and is no longer representative of the population.

Many models have been developed for censoring and truncation. This chapter focuses on the most frequently used model for censoring, the tobit model. Section 7.6 briefly reviews related models for truncation, multiple censoring, and sample selection.

7.1. The Problem of Censoring

Let y^* be a dependent variable that is *not* censored. Panel A of Figure 7.1 shows the distribution of y^*, where the height of the curve indicates the relative frequency of a given value of y^*. If we do not know the value of y^* when $y^* \leq 1$, corresponding to the shaded region, then y^* is a *latent* variable that cannot be observed over its entire range. The *censored* variable y is defined as

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 1 \\ 0 & \text{if } y^* \leq 1 \end{cases}$$

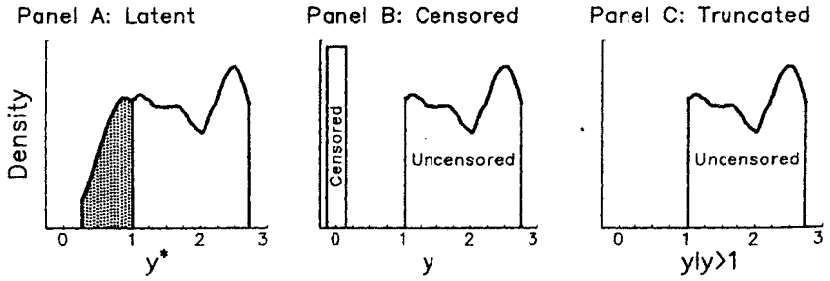

Figure 7.1. Latent, Censored, and Truncated Variables

Panel B plots the censored variable y with censored cases stacked at 0. The bar contains cases from the shaded region in panel A. Panel C plots the truncated variable $y \mid y > 1$ (i.e., y given that $y > 1$), which simply deletes the shaded region from panel A.

To see how censoring and truncation affect the LRM, consider the model $y^* = 1.2 + .08x + \varepsilon$, where all of the assumptions of the LRM apply, including the normality of the errors. Panel A of Figure 7.2 shows a sample of 200 with *no* censoring. The solid line is the OLS estimate $\widehat{y}^* = 1.18 + .08x$. If y^* were censored below at 1, we would know x for all observations, but observe y^* only for $y^* > 1$. In panel B, values of y^* at or below 1 are censored with $y = 0$ for censored cases. These are plotted with triangles. The three thick lines are the results of three approaches to estimation.

One way to estimate the parameters is with an OLS regression of y on x for all observations, with the censored data included as 0's. The resulting estimate $\widehat{y} = .95 + .11x$ is the long dashed line in panel B. The censored observations on the left pull down that end of the line, resulting in underestimates of the intercept and overestimates of the slope. This approach to censoring produces inconsistent estimates.

Since including censored observations causes problems, we might use OLS to estimate the regression after truncating the sample to exclude cases with a censored dependent variable. This changes the problem of censoring into the problem of a truncated sample. After deleting the cases at $y = 0$, the OLS estimate $\widehat{y} = 1.41 + .61x$ overestimates the intercept and underestimates the slope, as shown by the short dashed line. The uncensored observations at the left have pulled the line up, since those observations with large negative errors have been deleted. Truncation causes a correlation between x and ε which produces inconsistent estimates.

A third approach is to estimate the *tobit model*, sometimes referred to as the *censored regression model*. The tobit model uses all of the information, including information about the censoring, and provides consistent estimates of the parameters. ML estimates for the tobit model are shown by the solid line, which is indistinguishable from the estimates in panel A where there is no censoring.

Example of Censoring and Truncation: Prestige of the First Job

Chapter 2 used as an example the regression of the prestige of a scientist's first academic job. (See Table 2.1, p. 19, for a description of the

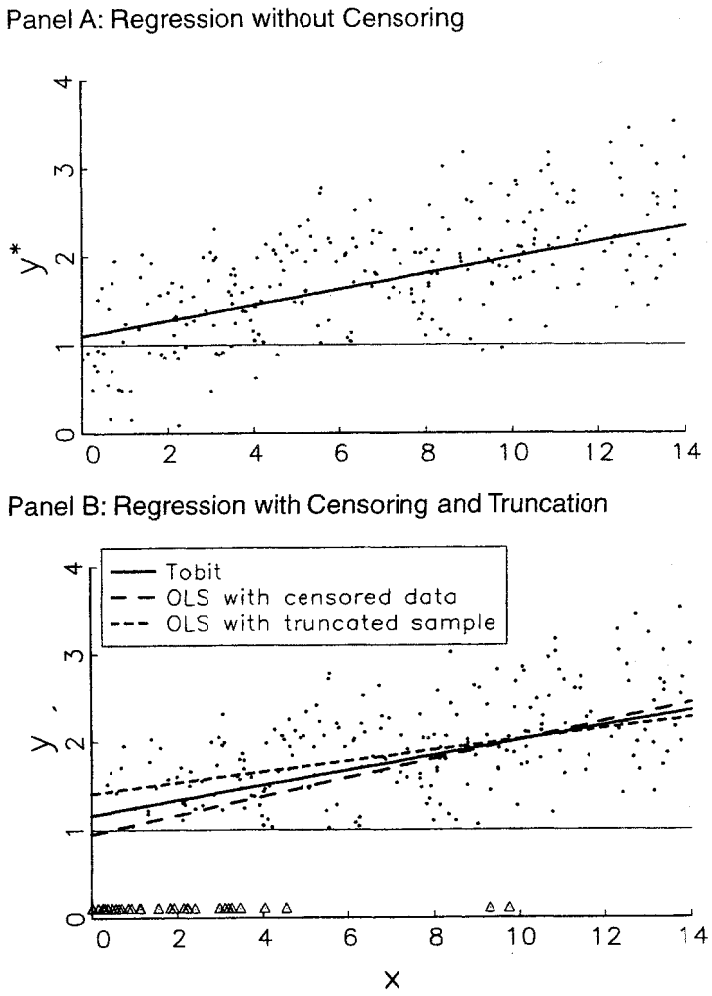

Figure 7.2. Linear Regression Model With and Without Censoring and Truncation

variables used.) The prestige of the job was unavailable for graduate programs rated below 1.0 and for departments without graduate programs. These cases were recoded to 1.0 and OLS was used to estimate the model. The estimates from Chapter 2 are reproduced in the column "OLS with Censored Data" in Table 7.1. Alternatively, we could truncate the sample by deleting the censored cases. The OLS estimates from the truncated sample are in the column "OLS with a Truncated Sample." Finally, tobit estimates are listed in the column "Tobit Analysis."

TABLE 7.1 Censoring and Truncation in the Analysis of the Prestige of the First Academic Job

Variable		*OLS with Censored Data*	*OLS with a Truncated Sample*	*Tobit Analysis*
Constant	β	1.067	1.413	0.685
	t/z	6.42	8.71	3.15
FEM	β	−0.139	0.101	−0.237
	β^{S_y}	−0.143	0.130	−0.194
	t/z	−1.54	1.19	−2.05
PHD	β	0.273	0.297	0.323
	β^{S}	0.267	0.354	0.252
	t/z	5.53	6.36	5.08
MENT	β	0.001	0.001	0.001
	β^{S}	0.080	0.069	0.072
	t/z	1.69	1.27	1.52
FEL	β	0.234	0.141	0.325
	β^{S_y}	0.240	0.180	0.267
	t/z	2.47	1.57	2.68
ART	β	0.023	0.006	0.034
	β^{S}	0.053	0.018	0.028
	t/z	0.79	0.24	0.93
CIT	β	0.004	0.002	0.005
	β^{S}	0.152	0.098	0.138
	t/z	2.28	1.27	2.06
	N	408	309	408
	R^2	0.210	0.201	

NOTE: β is an unstandardized coefficient; β^{S} is a fully standardized coefficient; β^{S_y} is a y-standardized coefficient; t/z is a t- or z-test of β.

The most important difference between the results of the tobit analysis and the two OLS regressions concerns the effect of gender. In the tobit analysis, the effect of being a woman is significant and negative. In the regression with censored data, the effect is substantially smaller and not significant. In the truncated regression, the effect is positive, although not significant. Thus, a key substantive result is dependent on the method of analysis. Other differences in relative magnitude and level of significance are also found.

7.2. Truncated and Censored Distributions

Before formally considering the tobit model, we need some results about truncated and censored normal distributions. These distributions are at the foundation of most models for truncation and censoring. Results are given for censoring and truncation on the left, which translates into censoring from below in the tobit model. Corresponding formulas for censoring and truncation on the right, and both on the left and on the right are available. For more details, see Johnson et al. (1994, pp. 156–162) or Maddala (1983, pp. 365–368).

7.2.1. The Normal Distribution

To indicate that y^* is distributed normally with mean μ and variance σ^2, we write $y^* \sim \mathcal{N}(\mu, \sigma^2)$. y^* has the pdf:

$$f(y^* \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y^* - \mu}{\sigma}\right)^2\right]$$

which is plotted in panel A of Figure 7.3. The cdf is

$$F(y^* \mid \mu, \sigma) = \int_{-\infty}^{y^*} f(z \mid \mu, \sigma)dz = \Pr(Y^* \leq y^*)$$

so that

$$\Pr(Y^* > y^*) = 1 - F(y^* \mid \mu, \sigma)$$

$F(\tau \mid \mu, \sigma)$ is the shaded region in panel A and $1 - F(\tau \mid \mu, \sigma)$ is the region to the right of τ.

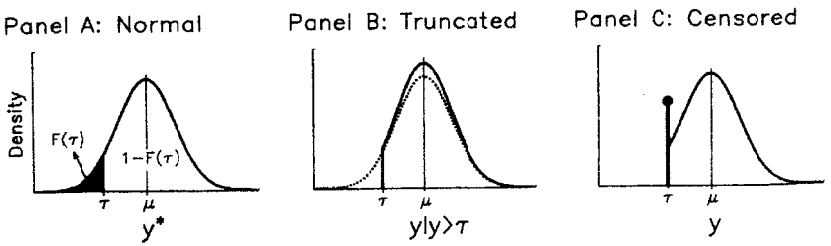

Figure 7.3. Normal Distribution With Truncation and Censoring

When $\mu = 0$ and $\sigma = 1$, the *standard normal distribution* is written in the simplified notation:

$$\phi(y^*) = f(y^* \mid \mu = 0, \sigma = 1)$$
$$\Phi(y^*) = F(y^* \mid \mu = 0, \sigma = 1)$$

Any normal distribution, regardless of its mean μ and variance σ^2, can be written as a function of the standard normal distribution. The pdf can be written as

$$f(y^* \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y^* - \mu}{\sigma}\right)^2\right] = \frac{1}{\sigma}\phi\left(\frac{y^* - \mu}{\sigma}\right) \quad [7.1]$$

and the cdf of y^* can be written as

$$\Pr(Y^* \leq y^*) = \Phi\left(\frac{y^* - \mu}{\sigma}\right) \quad [7.2]$$

So that,

$$\Pr(Y^* > y^*) = 1 - \Phi\left(\frac{y^* - \mu}{\sigma}\right)$$

Since the standard normal distribution is symmetric with a mean of 0, two identities follow that are frequently used to simplify other formulas:

$$\phi(\delta) = \phi(-\delta)$$
$$\Phi(\delta) = 1 - \Phi(-\delta)$$

These results are often used to simplify the equations in this chapter. For example, Equations 7.1 and 7.2 can be written as

$$f(y^* \mid \mu, \sigma) = \frac{1}{\sigma}\phi\left(\frac{\mu - y^*}{\sigma}\right)$$
$$\Pr(Y^* > y^*) = \Phi\left(\frac{\mu - y^*}{\sigma}\right)$$

7.2.2. The Truncated Normal Distribution

When values below τ are deleted, the variable $y \mid y > \tau$ has a truncated normal distribution. In terms of panel A of Figure 7.3, we want to consider the distribution of y^* in the unshaded region, while ignoring

all cases in the shaded region. The truncated pdf is created by dividing the pdf of the original distribution by the region to the right of τ. This forces the resulting distribution to have an area of 1:

$$f(y \mid y > \tau, \mu, \sigma) = \frac{f(y^* \mid \mu, \sigma)}{\Pr(Y^* > \tau)}$$

The truncated distribution is shown in panel B by the solid line. The mass of the shaded region has been distributed over the region to the right of τ, making the curve slightly higher over this region. This is seen by comparing the solid curve for the truncated distribution to the dotted line for the normal distribution without truncation. Using the results from Equations 7.1 and 7.2, we can write the truncated distribution as

$$f(y \mid y > \tau, \mu, \sigma) = \frac{\frac{1}{\sigma}\phi\left(\frac{y^* - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right)} = \frac{\frac{1}{\sigma}\phi\left(\frac{\mu - y^*}{\sigma}\right)}{\Phi\left(\frac{\mu - \tau}{\sigma}\right)}$$

Given that the left-hand side of the distribution has been truncated, $E(y \mid y > \tau)$ must be larger than $E(y^*) = \mu$. Specifically, if y^* is normal (Johnson et al., 1994, p. 156),

$$E(y \mid y > \tau) = \mu + \sigma\frac{\phi\left(\frac{\mu - \tau}{\sigma}\right)}{\Phi\left(\frac{\mu - \tau}{\sigma}\right)} = \mu + \sigma\,\lambda\left(\frac{\mu - \tau}{\sigma}\right) \qquad [7.3]$$

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

The inverse Mills ratio is used so frequently in this chapter that it deserves a careful examination. Figure 7.4 plots λ and its components ϕ and Φ as a function of $(\mu - \tau)/\sigma$. The quantity $(\mu - \tau)/\sigma$ is the number of standard deviations that the mean μ is above or below the truncation point. For example, $(\mu - \tau)/\sigma = 2$ means that μ is 2 standard deviations larger than τ. In Figure 7.4, assume that the mean μ is fixed, and consider the effects of changing τ. At the left of the figure, τ exceeds μ and truncation is more extreme. ϕ is larger than Φ, generating values of λ greater than 1. Moving to the right, τ decreases, the amount of truncation decreases, and $(\mu - \tau)/\sigma$ increases. With this change, Φ increases and is eventually larger than ϕ, resulting in smaller values of λ that eventually approach 0. Equation 7.3 shows that as λ approaches 0, the expected value of the truncated variable approaches μ. That is, as

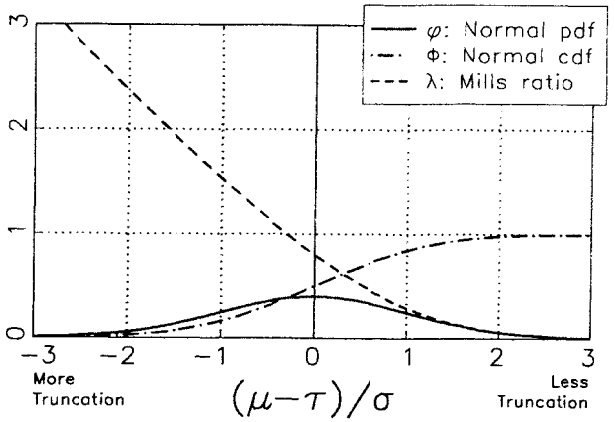

Figure 7.4. Inverse Mills Ratio

the area of truncation decreases, the effect of truncation on the mean approaches 0.

7.2.3. The Censored Normal Distribution

When a distribution is censored on the left, observations with values at or below τ are set to τ_y:

$$y = \begin{cases} y^* & \text{if } y^* > \tau \\ \tau_y & \text{if } y^* \leq \tau \end{cases}$$

Most often, $\tau_y = \tau$, but other values such as zero are also useful. Panel C of Figure 7.3 plots a censored normal variable, where the censored observations are indicated by the spike at $y = \tau$. From Equation 7.2, we know that if y^* is normal, then the probability of an observation being censored is

$$\Pr(\text{Censored}) = \Pr(y^* \leq \tau) = \Phi\left(\frac{\tau - \mu}{\sigma}\right)$$

and the probability of a case not being censored is

$$\Pr(\text{Uncensored}) = 1 - \Phi\left(\frac{\tau - \mu}{\sigma}\right) = \Phi\left(\frac{\mu - \tau}{\sigma}\right)$$

Thus, the expected value of a censored variable equals

$$\begin{aligned} E(y) &= [\Pr(\text{Uncensored}) \times E(y \mid y > \tau)] \\ &\quad + [\Pr(\text{Censored}) \times E(y \mid y = \tau_y)] \\ &= \left\{\Phi\left(\frac{\mu - \tau}{\sigma}\right)\left[\mu + \sigma\lambda\left(\frac{\mu - \tau}{\sigma}\right)\right]\right\} + \Phi\left(\frac{\tau - \mu}{\sigma}\right)\tau_y \end{aligned} \quad [7.4]$$

where the last equality uses Equation 7.3. Consider how the expected value of the censored value depends on τ. As τ approaches ∞, the probability of being censored approaches 1 and $E(y)$ approaches the censoring value τ_y. As τ approaches $-\infty$, the probability of being censored approaches 0 and $E(y)$ approaches the uncensored mean μ.

These results are now used to present the tobit model.

7.3. The Tobit Model for Censored Outcomes

For the tobit model, the structural equation is

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \quad [7.5]$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. The x's are observed for all cases. y^* is a latent variable that is observed for values greater than τ and is censored for values less than or equal to τ. The observed y is defined by the measurement equation:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > \tau \\ \tau_y & \text{if } y^* \leq \tau \end{cases} \quad [7.6]$$

Combining Equations 7.5 and 7.6,

$$y_i = \begin{cases} y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i & \text{if } y_i^* > \tau \\ \tau_y & \text{if } y_i^* \leq \tau \end{cases} \quad [7.7]$$

The tobit model can also be used in situations where there is censoring from above. For example, if incomes over \$100,000 were combined into the category "over \$100,000," income would be censored from above and the tobit model would be

$$y_i = \begin{cases} y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i & \text{if } y_i^* < \tau \\ \tau_y & \text{if } y_i^* \geq \tau \end{cases}$$

Results for censoring from above are given in Section 7.6.1.

In this section, I present the implications of censoring in several steps. First, I show the effects of independent variables on the probability of censoring. Next, I demonstrate the problems associated with using OLS with censored data or a truncated sample. These problems lead to the ML estimator for the tobit model. Finally, I consider several methods for interpreting the parameters in the tobit model. Before proceeding, I want to consider a potential source of confusion.

7.3.1. The Distinction Between τ and τ_y

Many authors assume that $\tau = \tau_y = 0$ or that $\tau = \tau_y$. This results in formulas that are simpler than mine. Unfortunately, this simplification can lead to confusion and incorrect results since in applications it is often the case that $\tau \neq \tau_y \neq 0$. Consequently, I make the distinction between τ and τ_y explicit. *The threshold τ determines whether y^* is censored. τ_y is the value assigned to y if y^* is censored.* While τ_y is often equal to τ, this is not always appropriate. Consider Tobin's original application. The cost of the cheapest durable good (i.e., τ) is not \$0, but for censored cases it is most reasonable to code $y = \tau_y = 0$ since these people did not purchase any goods. In my formulas, you can substitute $\tau = \tau_y = 0$ or $\tau = \tau_y$ to obtain formulas that match those in other sources. However, if you use formulas that equate these quantities, it is *essential* that these restrictions apply to your data.

7.3.2. The Distribution of Censoring

The probability of being censored depends on the proportion of the distribution of y^*, or, equivalently, ε, that falls below τ. The distribution of y given x is shown in panel A of Figure 7.5. $E(y^* \mid x)$ is the solid line with the distribution of y^* shown at three values of x. For example, at x_1 a vertical line is drawn with a normal curve coming out of the page. Censoring occurs when observations fall at or below the line $y^* = \tau$, which is indicated by the shaded region of the distribution. As the value of x increases, $E(y^* \mid x)$ increases, causing the proportion of the distribution that is censored to decrease. Thus, the region labeled A is larger than B, which is larger than C.

The probability of a case being censored for a given $\mathbf{x}$ is the region of the normal curve less than or equal to τ:

$$\Pr(\text{Censored} \mid \mathbf{x}_i) = \Pr(y_i^* \leq \tau \mid \mathbf{x}_i) = \Pr(\varepsilon_i \leq \tau - \mathbf{x}_i\boldsymbol{\beta} \mid \mathbf{x}_i)$$

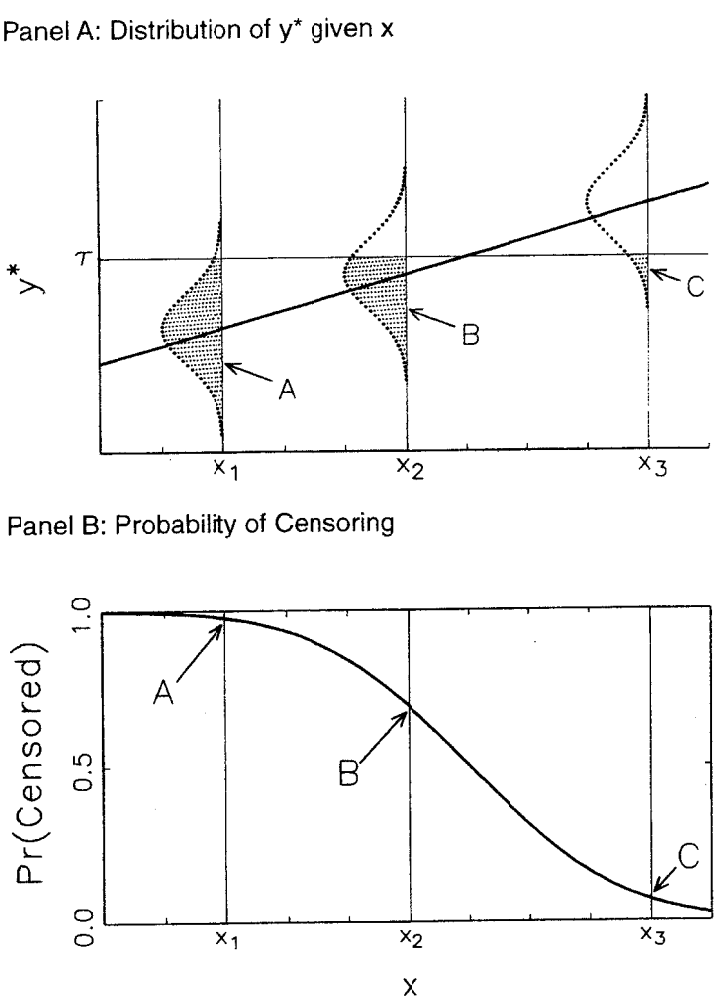

Figure 7.5. Probability of Being Censored in the Tobit Model

Since ε is distributed $\mathcal{N}(0, \sigma^2)$, ε/σ is distributed $\mathcal{N}(0, 1)$. Therefore,

$$\Pr(\text{Censored} \mid \mathbf{x}_i) = \Pr\left(\frac{\varepsilon_i}{\sigma} \leq \frac{\tau - \mathbf{x}_i\boldsymbol{\beta}}{\sigma} \,\middle|\, \mathbf{x}_i\right) = \Phi\left(\frac{\tau - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

and

$$\Pr(\text{Uncensored} \mid \mathbf{x}_i) = 1 - \Phi\left(\frac{\tau - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) = \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta} - \tau}{\sigma}\right)$$

To simplify the formulas that follow, let

$$\delta_i = \frac{\mathbf{x}_i\boldsymbol{\beta} - \tau}{\sigma}$$

δ is the number of standard deviations that $\mathbf{x}\boldsymbol{\beta}$ is above τ. (*How are* $\phi(\delta)$ *and* $\phi(-\delta)$ *related?* $\Phi(\delta)$ *and* $\Phi(-\delta)$?) Using this definition,

$$\Pr(\text{Censored} \mid \mathbf{x}_i) = \Phi(-\delta_i) \qquad [7.8]$$

$$\Pr(\text{Uncensored} \mid \mathbf{x}_i) = \Phi(\delta_i) \qquad [7.9]$$

Equation 7.8 is plotted in panel B of Figure 7.5. The points on the curve labeled *A*, *B*, and *C* correspond to the shaded regions in panel A. At the left, the change in Pr(Censored | x) is gradual as the thin tail moves over the threshold. The probability then decreases rapidly as the fat center of the curve passes over the threshold, and then changes slowly as the bottom tail passes over the threshold.

The Link Between Tobit and Probit

Deriving the probability of a case being censored is very similar to the derivation of the probability of an event in the probit model of Chapter 3. The structural models for probit and tobit are the same, but the measurement models differ. In the tobit model, we know the *value* of y^* when $y^* > \tau$, while in the probit model we only know *if* $y^* > \tau$. Since more information is available in tobit (i.e., we know y^* for same cases), estimates of the β's from tobit are more efficient than the estimates that would be obtained from a probit model. Further, since all cases are censored in probit, we have no way to estimate the variance of y^* and must assume that $\text{Var}(\varepsilon \mid \mathbf{x}) = 1$, while $\text{Var}(\varepsilon \mid \mathbf{x})$ can be estimated in the tobit model.

Example of the Probability of Censoring: Prestige of the First Job

The effects of doctoral prestige, gender, and having a postdoctoral fellowship on the probability that the prestige of the first job is censored are illustrated in Figure 7.6. The solid line with open squares shows the probability of censoring for women who were not fellows. Female fellows are less likely to have the prestige of their first job censored (i.e., to have a first job with prestige below 1), as shown by the solid line with

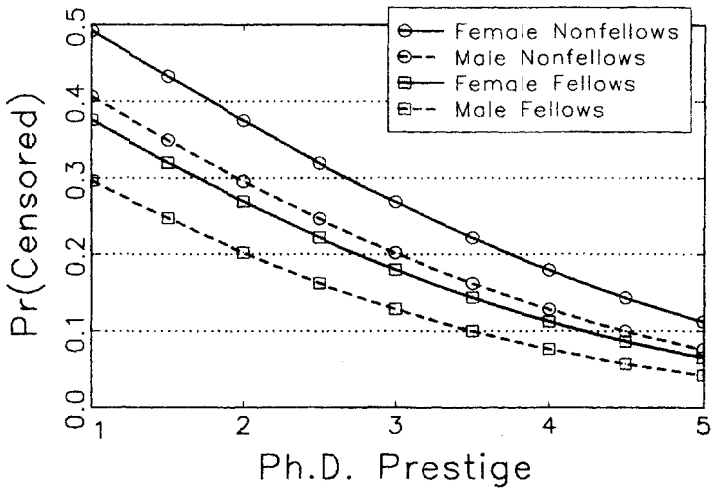

Figure 7.6. Probability of Being Censored by Gender, Fellowship Status, and Prestige of Doctoral Department

solid squares. When doctoral prestige is 1, female fellows have a probability of being censored of .38, which is .11 less than female nonfellows. When doctoral prestige is 5, female fellows have a probability of being censored of .07, which is .04 less than female nonfellows. Notice that the effect of being a fellow depends on doctoral prestige. The dashed lines show similar results for men. For male nonfellows, the probability decreases from .41 to .08, while for male fellows the probability decreases from .30 to .04. Being a female scientist increases the probability of being censored by .08 when doctoral prestige is 1 and .03 when doctoral prestige is 5.

These results suggest why our OLS results are biased in Table 7.1. Being a woman increases the probability of being censored, or, equivalently, of having a lower prestige job. When censored jobs were excluded from the analysis, the negative effect of being a woman on having a ranked job is not reflected in the sample. Consequently, the estimated effect of being a woman is positive, as shown in the column "OLS with a Truncated Sample" in Table 7.1. When unranked jobs are coded 1 and left in the sample, results are biased since these jobs are assigned a higher score than they would have had if the variable were not censored. That is, most of these cases would have had prestige scores lower than 1 if the data had been available. Consequently, the negative effect of being a

woman is underestimated, as shown in the column "OLS with Censored Data" in Table 7.1.

A more formal demonstration of the consequences of censoring for OLS estimation is now given.

7.3.3. Problems Introduced by Censoring

Being unable to observe y^* over its entire range causes problems for the LRM. Most immediately, a decision must be made on how to handle the censored observations. There are two approaches that were used frequently prior to the acceptance of the tobit model:

1. A truncated sample is created by deleting cases where the dependent variable is censored. The model is estimated with OLS using the truncated sample.
2. A censored dependent variable is created in which all censored observations are assigned the value τ_y. The model is estimated with OLS using the censored dependent variable.

Berndt (1991, pp. 614–617) provides an interesting analysis of the consequences of using these approaches in research on the labor supply. Here, I demonstrate the problems with these approaches, and in the process present results that are useful for interpreting the tobit model.

Analyzing a Truncated Sample

The structural model for the latent variable is $y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon$. Since $E(\varepsilon \mid \mathbf{x}) = 0$, $E(y^* \mid \mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$. With truncation, our model is

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \quad \text{for all } i \text{ such that } y_i > \tau$$

The dependent variable is the truncated variable $y \mid y > \tau$. Taking expectations,

$$\begin{aligned} E(y_i \mid y_i > \tau, \mathbf{x}_i) &= E(\mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \mid y_i > \tau, \mathbf{x}_i) \\ &= \mathbf{x}_i\boldsymbol{\beta} + E(\varepsilon_i \mid y_i > \tau, \mathbf{x}_i) \end{aligned}$$

If $E(\varepsilon \mid y > \tau, \mathbf{x}) = 0$, then $E(y \mid y > \tau, \mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$ and the model remains linear, which would justify OLS estimation. However, $E(\varepsilon \mid y > \tau, \mathbf{x})$ is not zero. From Equation 7.3, it follows that

$$E(y_i \mid y_i > \tau, \mathbf{x}_i) = \mathbf{x}_i\boldsymbol{\beta} + \sigma\,\lambda(\delta_i) \qquad [7.10]$$

where σ is the standard deviation of ε, $\delta = (\mathbf{x}\boldsymbol{\beta} - \tau)/\sigma$, and λ is the inverse Mills ratio.

Figure 7.7 illustrates the effects of truncation. If there were no truncation or censoring, the sample would correspond to the dots, both above and below the truncation point at $y^* = \tau$, with the OLS estimate of $E(y^* \mid x)$ shown by the solid line. Truncation occurs at $\tau = 2$, which is indicated by a horizontal line. With truncation, all observations at or below τ are dropped from the analysis. The relationship between x and the truncated expectation $E(y \mid y > \tau, x)$ is shown by the long dashed curve. This is the top curve. At the right, $E(y \mid y > \tau, x)$ is indistinguishable from $E(y^* \mid x)$ since so few cases are truncated. As we move to the left, $E(y \mid y > \tau, x)$ moves above $E(y^* \mid x)$ since smaller values of y^* have been excluded from the sample. As x continues to move to the left, $E(y \mid y > \tau, x)$ becomes closer and closer to τ. Given the difference between $E(y^* \mid x)$ and $E(y \mid y > \tau, x)$, it is clear why OLS produces inconsistent estimates when the sample is truncated.

Another way of thinking about the problems introduced by truncation is to consider the regression model implied by Equation 7.10:

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \sigma\lambda_i + e_i \qquad [7.11]$$

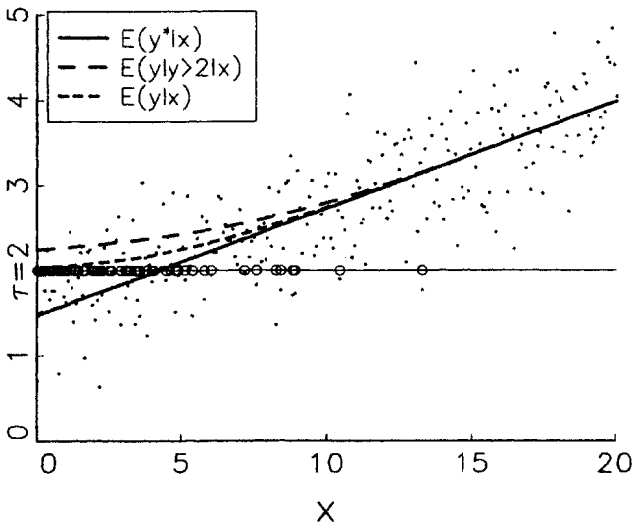

Figure 7.7. Expected Values of y^*, $y \mid y > \tau$, and y in the Tobit Model

where λ_i is used for $\lambda(\delta_i)$ to emphasize that λ_i may be thought of as another variable in the equation and σ can be thought of as the slope coefficient for the variable λ. If we estimate $\boldsymbol{\beta}$ using $y = \mathbf{x}\boldsymbol{\beta} + \varepsilon$, we have a misspecified model that excludes λ. The OLS estimates will be inconsistent.

Analyzing Censored Data

A second approach is to analyze the entire sample after assigning $y = \tau_y$ to censored cases. In Figure 7.7, the censored observations are indicated by the circles located on the line $\tau = 2$. Since values of y^* below 2 have been set equal to 2, $E(y \mid x)$ is above $E(y^* \mid x)$ as shown by the short dashed line. This line is below $E(y \mid y > \tau, x)$ since the censored cases have not been eliminated, but only given an unrealistically large value. If we use OLS to estimate a regression using the entire sample after assigning τ_y to censored observations, the estimates are inconsistent.

More formally, with censoring our model becomes

$$y_i = \begin{cases} y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i & \text{if } y_i^* > \tau \\ \tau_y & \text{if } y_i^* \leq \tau \end{cases} \qquad [7.12]$$

Applying Equation 7.4, the expected value of y given $\mathbf{x}$ is the sum of components for uncensored and censored cases:

$$E(y_i \mid \mathbf{x}_i) = [\Pr(\text{Uncensored} \mid \mathbf{x}_i) \times E(y_i \mid y_i > \tau, \mathbf{x}_i)] \qquad [7.13]$$
$$+ [\Pr(\text{Censored} \mid \mathbf{x}_i) \times \tau_y]$$

Using Equations 7.8 and 7.9 with $\delta = (\mathbf{x}\boldsymbol{\beta} - \tau)/\sigma$,

$$E(y_i \mid \mathbf{x}_i) = [\Phi(\delta_i) \times E(y_i \mid y_i > \tau, \mathbf{x}_i)] + [\Phi(-\delta_i) \times \tau_y] \qquad [7.14]$$

Substituting results from Equations 7.10 and 7.12, and simplifying,

$$E(y_i \mid \mathbf{x}_i) = \Phi(\delta_i)\mathbf{x}_i\boldsymbol{\beta} + \sigma\phi(\delta_i) + \Phi(-\delta_i)\tau_y \qquad [7.15]$$

$E(y \mid \mathbf{x})$ is nonlinear in $\mathbf{x}$, so that estimating the regression of y on $\mathbf{x}$ results in inconsistent estimates of the parameters for the regression of y^* on $\mathbf{x}$. (*What happens to Equation 7.15 if* $\Phi(\delta) = 1$? *If* $\Phi(\delta) = 0$?)

7.4. Estimation

In the presence of censoring, OLS is inconsistent. One approach to estimating the tobit model is based on Equation 7.11: $y = \mathbf{x}\boldsymbol{\beta} + \sigma\lambda + e$. Heckman (1976) proposed a two-stage estimator in which λ is estimated by probit in the first stage, and $y = \mathbf{x}\boldsymbol{\beta} + \sigma\hat{\lambda} + e$ is estimated by OLS in the second stage. Since this estimator is less efficient and no easier to compute than the ML estimator, I do not consider it further.

ML estimation for the tobit model involves dividing the observations into two sets. The first set contains uncensored observations, which ML treats in the same way as the LRM. The second set contains censored observations. For these observations, we do not know the specific value of y^*, but can proceed by computing the probability of being censored and using this quantity in the likelihood equation. Figure 7.8 illustrates the approach used for three observations represented by solid circles. At each value of x, there is a normal curve showing the distribution of y^* given x. For uncensored observations, the distance from the observation to the normal curve is the likelihood of that observation for a given $\boldsymbol{\beta}$ and σ. The line at $y^* = \tau$ indicates where censoring occurs. For the censored observations, such as (x_1, y_1^*), we do not know the value of y^* and hence cannot use the height of the normal curve at that point for the likelihood. Since all we know for censored cases is that $y^* \leq \tau$, we

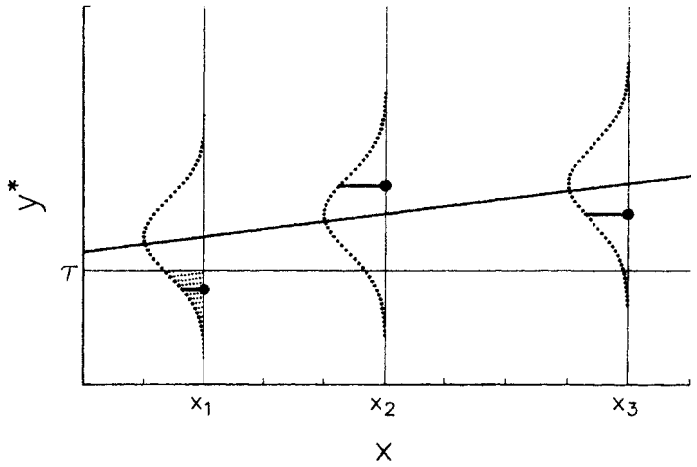

Figure 7.8. Maximum Likelihood Estimation for the Tobit Model

use the probability of being censored as the likelihood. This is indicated by the shaded region.

Formally, for uncensored observations,

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \quad \text{for } y^* > \tau$$

where $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$. As in Equation 2.8, the log likelihood equation for uncensored observations is

$$\ln L_U(\boldsymbol{\beta}, \sigma^2) = \sum_{\text{Uncensored}} \ln \frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

In Figure 7.8, $\ln L_U$ is the sum of the logs of the spikes at (x_2, y_2^*) and (x_3, y_3^*).

For censored observations, we know $\mathbf{x}$ and that $y^* \leq \tau$, so we can compute

$$\Pr(y_i^* \leq \tau \mid \mathbf{x}_i) = \Phi\left(\frac{\tau - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) \qquad [7.16]$$

Thus, for the first observation in Figure 7.8, we are computing the area of the shaded region at or below $y = \tau$ rather than the height of the pdf at y^*. Using Equation 7.16, we can write that part of the likelihood function that applies to censored observations as

$$L_C(\boldsymbol{\beta}, \sigma^2) = \prod_{\text{Censored}} \Phi\left(\frac{\tau - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

Taking logs,

$$\ln L_C(\boldsymbol{\beta}, \sigma^2) = \sum_{\text{Censored}} \ln \Phi\left(\frac{\tau - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

Combining the results for censored and uncensored observations,

$$\ln L(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{y}, \mathbf{X}) = \sum_{\text{Uncensored}} \ln \frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) + \sum_{\text{Censored}} \ln \Phi\left(\frac{\tau - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

While this likelihood equation is unusual with its combination of the pdf for uncensored observations and the cdf for censored observations,

Amemiya (1973) shows that if the assumptions of the tobit model hold, the usual properties of ML will apply.

Many programs estimate the tobit model by ML, including LIMDEP, Markov, Stata, and SAS's LIFEREG. Each program requires a method for specifying which observations are censored. Most commonly, you specify the value of τ and the program assumes that each observation with the dependent variable less than or equal to τ is censored.

7.4.1. Violations of Assumptions

The ML estimator for the tobit model assumes that the errors are normal and homoscedastic. In the LRM, if these assumptions are violated the estimates remain consistent, but not efficient. This is not the case in the tobit model.

Heteroscedasticity. Maddala and Nelson (1975) show that the ML estimator for the tobit model is inconsistent if there is heteroscedasticity. Maddala (1983, pp. 179–182) illustrates the effects of heteroscedasticity in estimates for several models, and Arabmazar and Schmidt (1981) provide further analysis of the robustness of the ML estimator to heteroscedasticity. The log likelihood equation can be modified to account for heteroscedasticity by replacing σ by σ_i. LIMDEP provides ML estimates when heteroscedasticity is of the form: $\sigma_i = \sigma \exp(\mathbf{z}_i\boldsymbol{\gamma})$. See Greene (1993, pp. 698–700) for further discussion.

Nonnormal Errors. The ML estimator is inconsistent when the errors are nonnormal (Arabmazar & Schmidt, 1982). Estimation of the tobit model with nonnormal errors is possible using programs for event history analysis such as SAS's LIFEREG or LIMDEP. The link between tobit analysis and event history analysis is considered further in Chapter 9.

7.5. Interpretation

There are three outcomes that can be of interest in the tobit model: (1) the latent variable y^*; (2) the truncated variable $y \mid y > \tau$; and (3) the censored variable y. This section presents methods for interpreting changes in the expected values of each of these outcomes using partial and discrete change. Since $y \mid y > \tau$ and y are rarely used except in economics, they are discussed only briefly.

7.5.1. Change in the Latent Outcome

In many applications, changes in the latent y^* are of primary interest. Tobit analysis provides consistent estimates of the effects of the independent variables on the latent y^*. The expected value of y^* is

$$E(y^* \mid \mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$$

and the partial derivative with respect to x_k is

$$\frac{\partial E(y^* \mid \mathbf{x})}{\partial x_k} = \beta_k$$

For a continuous independent variable x_k, we can state:

- For a unit increase in x_k, there is an expected change of β_k units in y^*, holding all other variables constant.

For a dichotomous variable,

- Having characteristic x_k (as opposed to not having the characteristic) increases the expected value of y^* by β_k units, holding all other variables constant.

Since the model is linear in y^*, the effect of x_k does not depend on the value of x_k or the values of the other x's.

Standardized Coefficients

Following the arguments presented in Section 2.2.1 (p. 15) for the LRM, fully standardized and semi-standardized coefficients can be computed as

$$\beta_k^{S_x} = \sigma_k \beta_k, \qquad \beta_k^{S_{y^*}} = \frac{\beta_k}{\sigma_{y^*}} \qquad \beta_k^{S} = \frac{\sigma_k \beta_k}{\sigma_{y^*}}$$

where σ_{y^*} is the *unconditional* standard deviation of y^* and σ_k is the standard deviation of x_k. Since y^* is a latent variable, σ_{y^*} cannot be computed directly from the observed data. To deal with this problem, Roncek (1992) suggested an "analogue to a standardized coefficient" that uses the standard deviation of y^* *conditional* on $\bar{\mathbf{x}}$: $\beta_k^{S^*} = \beta_k \sigma_k / \sigma_{y^* \mid \bar{\mathbf{x}}}$. Since $\sigma_{y^* \mid \bar{\mathbf{x}}}$ and σ_{y^*} can be quite different (*Why?*), Roncek's approximation should not be used. Instead, the unconditional variance of y^* should be computed with the quadratic form:

$$\widehat{\sigma}^2_{y^*} = \widehat{\boldsymbol{\beta}}' \widehat{\mathrm{Var}}(\mathbf{x}) \widehat{\boldsymbol{\beta}} + \widehat{\sigma}^2_{\varepsilon}$$

where $\widehat{\text{Var}}(\mathbf{x})$ is the estimated covariance matrix among the x's and $\hat{\sigma}_\varepsilon^2$ is the ML estimate of the variance of ε.

Example of Partial Change in y: Prestige of the First Job*

The *tobit* coefficients in Table 7.1 can be interpreted in the same way as the results of the LRM in Chapter 2. To illustrate this, consider the effects of *FEM* and *PHD*.

- Being a female scientist decreases the expected prestige of the first job by .24 points on a five-point scale, holding all other variables constant. Further, being a female scientist decreases the expected prestige of the first job by .19 standard deviations, holding all other variables constant.
- For a unit increase in the prestige of the doctoral department, the prestige of the first job is expected to increase by .32 units, holding all other variables constant. For a standard deviation increase in the prestige of the doctoral department, the prestige of the first job is expected to increase by .25 standard deviations, holding all other variables constant.

The effects of being female and doctoral prestige are significant at the .01 level for one-tailed tests.

7.5.2. Change in the Truncated Outcome

The truncated variable $y \mid y > \tau$ is defined only for those observations that are not truncated. If the dependent variable is expenditures on durable goods, the truncated outcome is how much was spent by people who purchased durable goods. Those who did not purchase goods are excluded from the sample. In economics, this outcome may be of considerable interest. For example, manufacturers of durable goods might want to know how much money consumers will spend and may be uninterested in how much consumers would have spent if durable goods could have been purchased for less than the threshold τ. Whether the truncated outcome is of interest depends on the substantive focus of the research.

Earlier, we showed that the expected value of the truncated outcome is

$$E(y \mid y > \tau, \mathbf{x}) = \mathbf{x}\boldsymbol{\beta} + \sigma\,\lambda(\delta) \qquad [7.17]$$

where $\lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)$ and $\delta = (\mathbf{x}\boldsymbol{\beta} - \tau)/\sigma$. The expected value is nonlinear in the x's, and, consequently, β_k cannot be interpreted as the

effect of a unit increase in x_k on the expected value of the truncated outcome. The partial derivative of $E(y \mid y > \tau, \mathbf{x})$ with respect to x_k is

$$\frac{\partial E(y \mid y > \tau)}{\partial x_k} = \beta_k[1 - \delta\lambda(\delta) - \lambda(\delta)^2] \qquad [7.18]$$

Greene (1993, p. 688) shows that the quantity in square brackets must fall between 0 and 1, approaching 1 as $\mathbf{x}\boldsymbol{\beta}$ increases. Thus, $\partial E(y \mid y > \tau, \mathbf{x})/\partial x_k$ approaches $\partial E(y^* \mid \mathbf{x})/\partial x_k$ as $\mathbf{x}\boldsymbol{\beta}$ increases. This is seen by comparing the solid and long dashed lines in Figure 7.7. For a dichotomous variable, the discrete change as the variable moves from 0 to 1 should be used instead of the partial derivative:

$$\frac{\Delta E(y \mid y > \tau, \mathbf{x})}{\Delta x_k} = E(y \mid y > \tau, \mathbf{x}, x_k = 1) - E(y \mid y > \tau, \mathbf{x}, x_k = 0)$$

For both the partial and the discrete change, the change depends on the level of all of the x's in the model. As a summary measure, the partial or discrete change for each x_k with all other variables held at their means is often used.

7.5.3. Change in the Censored Outcome

The censored outcome y equals the latent y^* when the dependent variable is observed, and equals τ_y (usually τ or 0) when the dependent variable is censored. If the dependent variable is expenditures on durable goods, it is useful to let $\tau_y = 0$. Then $E(y \mid \mathbf{x})$ is the expected actual expenditures of those with a given $\mathbf{x}$. Those who are censored on y^* are included as 0's, which is how much they actually spent.

From Equation 7.15,

$$E(y \mid \mathbf{x}) = \Phi(\delta)\mathbf{x}\boldsymbol{\beta} + \sigma\,\phi(\delta) + \Phi(-\delta)\tau_y$$

where $\delta = (\mathbf{x}\boldsymbol{\beta} - \tau)/\sigma$. The partial derivative with respect to x_k is

$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = \Phi(\delta)\beta_k + (\tau - \tau_y)\,\phi(\delta)\frac{\beta_k}{\sigma}$$

If $\tau_y = \tau$, then we obtain the simpler result:

$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = \Phi(\delta)\beta_k = \Pr(\text{Uncensored} \mid \mathbf{x})\beta_k \qquad [7.19]$$

Regardless of the value of τ_y, as the probability of a case being censored approaches 0, the partial derivative of y approaches the partial for y^*. This is illustrated by comparing the solid and short dashed lines in Figure 7.7. For a dichotomous variable, the discrete change as the variable moves from 0 to 1 should be used instead of the partial derivative:

$$\frac{\Delta E(y \mid \mathbf{x})}{\Delta x_k} = E(y \mid \mathbf{x}, x_k = 1) - E(y \mid \mathbf{x}, x_k = 0)$$

For both the partial and the discrete change, the change depends on the level of all of the x's in the model. As a summary measure, the partial or discrete change for each x_k with all other variables held at their means is often used.

7.5.4. McDonald and Moffitt's Decomposition

McDonald and Moffitt (1980) suggest a decomposition of $\partial E(y)/\partial x_k$ that highlights two sources of change in the censored outcome. The simplest way to derive their decomposition is to differentiate Equation 7.13 by parts and apply the product rule. After a great deal of algebra,

$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = \Pr(U \mid \mathbf{x}) \frac{\partial E(y \mid y > \tau, \mathbf{x})}{\partial x_k} + [E(y \mid y > \tau, \mathbf{x}) - \tau_y] \frac{\partial \Pr(U \mid \mathbf{x})}{\partial x_k}$$

where $\Pr(U \mid \mathbf{x})$ is the probability of an observation being uncensored given $\mathbf{x}$. When $\tau_y = 0$, this results in the more commonly found version of this decomposition. It is important to realize that if τ_y is not 0, the simpler formula is not appropriate.[1]

The decomposition shows that when x_k changes, it affects the expectation of y^* for uncensored cases weighted by the probability of being uncensored, and it affects the probability of being uncensored weighted by the expected value for uncensored cases minus the censoring value τ_y. While this decomposition is useful for understanding how changes in the observed cases occur, its application depends on one's interest in y as opposed to y^*.

[1] Roncek (1992) provides a detailed discussion of the McDonald–Moffitt decomposition. While his formulas assume $\tau = \tau_y = 0$, his example has $\tau = \tau_y \neq 0$.

7.6. Extensions

The tobit model has been extended in many ways. Amemiya (1985, pp. 360–411), Berndt (1991, pp. 716–649), Breen (1996), and Maddala (1983, pp. 149–290) discuss many of these extensions, most of which can be estimated with LIMDEP (Greene, 1995, Chapter 27). In this section, I consider several basic extensions. The discussion is by no means comprehensive.

7.6.1. Upper Censoring

The simplest extension of the tobit model with censoring from below is the tobit model with censoring from above:

$$y = \begin{cases} \mathbf{x}\boldsymbol{\beta} + \varepsilon & \text{if } y^* < \tau \\ \tau_y & \text{if } y^* \geq \tau \end{cases}$$

This model can be obtained from the model with lower censoring simply by changing the sign of y. Censoring y from above at τ is identical to censoring $-y$ from below at $-\tau$. Since this simple change has subtle effects on the signs in many formulas, I present key results here. The probability of censoring is

$$\Pr(\text{Censored} \mid \mathbf{x}) = \Phi(\delta)$$

where $\delta = (\mathbf{x}\boldsymbol{\beta} - \tau)/\sigma$. Expected values are

$$E(y^* \mid \mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$$
$$E(y \mid y < \tau, \mathbf{x}) = \mathbf{x}\boldsymbol{\beta} - \sigma\,\lambda(-\delta)$$
$$E(y \mid \mathbf{x}) = \Phi(-\delta)\mathbf{x}\boldsymbol{\beta} - \sigma\,\phi(\delta) + \Phi(\delta)\tau_y$$

The partial derivatives with respect to x_k are

$$\frac{\partial E(y^*)}{\partial x_k} = \beta_k$$
$$\frac{\partial E(y \mid y < \tau)}{\partial x_k} = \beta_k[1 + \delta\lambda(-\delta) - \lambda(-\delta)^2]$$
$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = \Phi(-\delta)\beta_k + (\tau_y - \tau)\phi(\delta)\frac{\beta_k}{\sigma}$$

7.6.2. Upper and Lower Censoring

Rosett and Nelson (1975) developed the *two-limit tobit model* to allow both upper and lower censoring. With upper and lower censoring,

$$y = \begin{cases} \tau_L & \text{if } y^* \le \tau_L \\ y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon_i & \text{if } \tau_L < y^* < \tau_U \\ \tau_U & \text{if } y^* \ge \tau_U \end{cases}$$

A common application of this model is when the outcome is a probability or a percentage. For example, Saltzman (1987) examines the effects of political action committee contributions on the voting of members of the House of Representatives on labor issues. The dependent variable is the percentage of times that a member of the House voted for issues benefiting labor. Saltzman argues that the dependent variable is truncated at 100 since some representatives that voted pro-labor 100% of the time would probably have voted for strongly pro-labor bills that were never introduced for a vote. Thus, they were more positive than indicated by the 100% vote. The same logic would apply to those who never voted pro-labor. Similar reasoning was used by Fronstin and Holtmann (1994) in studying the percentage of houses in a development that were damaged by Hurricane Andrew, and by Sullivan and Worden (1990) in their study of the probability that an individual will file for bankruptcy.

With two limits, the likelihood function includes components for upper censoring, lower censoring, and no censoring. Defining $\delta_L = (\tau_L - \mathbf{x}\boldsymbol{\beta})/\sigma$ and $\delta_U = (\tau_U - \mathbf{x}\boldsymbol{\beta})/\sigma$, it can be shown that

$$\Pr(y = \tau_L \mid \mathbf{x}_i) = \Phi(\delta_L)$$
$$\Pr(y = \tau_U \mid \mathbf{x}_i) = 1 - \Phi(\delta_U) = \Phi(-\delta_U)$$

Then

$$\ln L = \sum_{\text{Lower}} \ln \Phi\left(\frac{\tau_L - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right) + \sum_{\text{Uncensored}} \ln \frac{1}{\sigma}\phi\left(\frac{y - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) + \sum_{\text{Upper}} \ln \Phi\left(\frac{\mathbf{x}\boldsymbol{\beta} - \tau_U}{\sigma}\right)$$

Interpretation proceeds along the lines used for the single-limit tobit model. For the latent outcome,

$$E(y^* \mid \mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$$

such that

$$\frac{\partial E(y^* \mid \mathbf{x})}{\partial x_k} = \frac{\Delta E(y^* \mid \mathbf{x})}{\Delta x_k} = \beta_k$$

The formulas for truncated and censored outcomes are generalizations of those discussed above. The expected value for the truncated outcome is (Maddala, 1983, pp. 160–162)

$$E(y \mid \tau_U > y > \tau_L, \mathbf{x}) = \mathbf{x}\boldsymbol{\beta} + \sigma \frac{\phi(\delta_L) - \phi(\delta_U)}{\Phi(\delta_U) - \Phi(\delta_L)}$$

The partial with respect to x_k is

$$\frac{\partial E(y \mid \tau_U > y > \tau_L, \mathbf{x})}{\partial x_k} = \beta_k\left(1 + \frac{\delta_L\phi(\delta_L) - \delta_U\phi(\delta_U)}{\Phi(\delta_U) - \Phi(\delta_L)} - \left[\frac{\phi(\delta_L) - \phi(\delta_U)}{\Phi(\delta_U) - \Phi(\delta_L)}\right]^2\right)$$

If x_k is dichotomous,

$$\frac{\Delta E(y \mid \tau_U > y > \tau_L, \mathbf{x})}{\Delta x_k} = E(y \mid \tau_U > y > \tau_L, \mathbf{x}, x_k = 1) - E(y \mid \tau_U > y > \tau_L, \mathbf{x}, x_k = 0)$$

For the observed outcome:

$$\begin{aligned} E(y \mid \mathbf{x}) &= [\tau_L \times \Pr(y = \tau_L \mid \mathbf{x}_i)] + [\tau_U \times \Pr(y_i = \tau_U \mid \mathbf{x}_i)] \\ &\quad + [E(y \mid \tau_L < y^* < \tau_U, \mathbf{x}) \times \Pr(\tau_L < y^* < \tau_U \mid \mathbf{x}_i)] \\ &= \tau_L\Phi(\delta_L) + \tau_U\Phi(-\delta_U) \\ &\quad + [\Phi(\delta_U) - \Phi(\delta_L)]\left[\mathbf{x}\boldsymbol{\beta} + \sigma \frac{\phi(\delta_L) - \phi(\delta_U)}{\Phi(\delta_U) - \Phi(\delta_L)}\right] \end{aligned}$$

Differentiating results in the simple expression:

$$\frac{\partial E(y \mid \mathbf{x})}{\partial x_k} = [\Phi(\delta_U) - \Phi(\delta_L)]\beta_k = \Pr(\text{Uncensored} \mid \mathbf{x})\beta_k$$

If x_k is a dichotomous variable,

$$\frac{\Delta E(y \mid \mathbf{x})}{\Delta x_k} = E(y \mid \mathbf{x}, x_k = 1) - E(y \mid \mathbf{x}, x_k = 0)$$

7.6.3. The Truncated Regression Model

Truncation occurs when no information for the dependent or independent variables is available for cases where the dependent variable is above or below a given level. For example, if you sample only individuals with incomes above $100,000, you have a sample that is truncated from below. The *truncated regression model* is used to analyze these types of data. The structural model is

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i \quad \text{for all } i \text{ such that } y_i < \tau$$

This corresponds to the first part of the structural equation for the tobit model with censoring from above. The likelihood of each observation is the same as for uncensored observations in the tobit model, except that the likelihood must be adjusted by the area of the normal distribution that has been truncated:

$$f(y_i) = \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{\tau - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}$$

The log likelihood function becomes $\ln L = \sum_i \ln f(y_i)$. The expected value $E(y \mid y < \tau, \mathbf{x})$ and the partial derivatives are the same as for the tobit model.

The importance of taking truncation into account is illustrated with results from Hausman and Wise's (1977) analysis of the New Jersey Negative Income Tax Experiment. In this study, the sample was truncated to exclude families with incomes more than 1.5 times the poverty level. Table 7.2 presents OLS estimates that ignore truncation and ML estimates of the truncated regression model. The ML estimates are as much as 5.4 times larger, with coefficients often having larger z-values. These results clearly illustrate how large the bias introduced by OLS estimation can be.

7.6.4. Individually Varying Limits

The tobit model can be generalized to allow censoring limits that differ from individual to individual. This extension is closely related to event history analysis and is discussed in Section 9.4.

TABLE 7.2 Hausman and Wise's OLS and ML Estimates From a Sample With Truncation

Variable		*OLS Estimates*	*ML Estimates*	*Ratio ML/OLS*
Constant	β	8.203	9.102	1.11
	t/z	90.14	356.95	3.96
Education	β	0.010	0.015	1.54
	t/z	1.67	2.09	1.25
IQ	β	0.002	0.006	3.81
	t/z	1.00	1.27	1.27
Training	β	0.002	0.007	2.95
	t/z	1.38	2.10	1.52
Union	β	0.090	0.246	2.74
	t/z	2.95	2.78	0.94
Illness	β	−0.076	−0.226	2.97
	t/z	−2.01	−2.11	1.05
Age	β	−0.003	−0.016	5.40
	t/z	−1.67	−3.06	1.83

NOTE: N=684. β is an unstandardized coefficient. t/z is a z-test of β for ML and a t-test of β for OLS.

7.6.5. Models for Sample Selection

Sample selection models generalize the tobit and truncated regression models by explicitly modeling the mechanism that selects observations as being censored or uncensored. There is a vast and growing literature on sample selection models (see Amemiya, 1985, Chapter 10; Maddala, 1983, Chapters 7–8; Manski, 1995, for further details). Here, I consider only the simplest model for sample selection, sometimes known as the Heckman model (1976).

In the tobit and truncated regression models, the structural model, is

$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$$

Instead of y^* being observed when $y^* > \tau$, assume that y^* is observed based on the value of a second latent variable z^*, where

$$z_i^* = \mathbf{w}_i\boldsymbol{\alpha} + \nu_i \qquad [7.20]$$

$\mathbf{x}$ and $\mathbf{w}$ can have variables in common. y^* is observed only when $z^* > 0$. To estimate the model, we assume that the errors are normally distributed such that

$$\begin{pmatrix}\varepsilon_i \\ \nu_i\end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}\sigma_\varepsilon^2 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & 1\end{pmatrix}\right]$$

where ρ is the correlation between ε and ν, and $\text{Var}(\nu)$ is assumed to be 1 to identify the model. Since ν is distributed normally, Equation 7.20 specifies a probit model where $z = 1$ (and y^* is observed) if $z^* > 0$.

Using derivations similar to those for the tobit model (Greene, 1993, pp. 709–711) results in the expected value of the observed y:

$$E(y_i \mid z_i = 1) = \mathbf{x}_i\boldsymbol{\beta} + \gamma\frac{\phi(-\mathbf{w}_i\boldsymbol{\alpha})}{\Phi(-\mathbf{w}_i\boldsymbol{\alpha})} = \mathbf{x}_i\boldsymbol{\beta} + \gamma\lambda_i$$

which is very similar to Equation 7.10. Regressing y on $\mathbf{x}$ for observations where $z = 1$ would produce inconsistent estimates since λ has been excluded. Heckman's two-step estimation involves first estimating the probit model in Equation 7.20 and computing

$$\widehat{\lambda}_i = \frac{\phi(-\mathbf{w}_i\widehat{\boldsymbol{\alpha}})}{\Phi(-\mathbf{w}_i\widehat{\boldsymbol{\alpha}})}$$

and then estimating the regression of y on $\mathbf{x}$ and $\widehat{\lambda}$.

7.7. Conclusions

This chapter has only touched on the rich set of models that deal with censoring, truncation, and sample selection. In all of these models, the basic problem is the same. Due to some data collection mechanism, data are missing on some of the observations in a systematic way. As a consequence, the LRM provides biased and inconsistent estimates.

7.8. Bibliographic Notes

While censored and truncated distributions have a long history in biometrics, engineering, and statistics, within the social sciences structural models for censoring and truncation originated with Tobin's (1958) article on household expenditures for durable goods. Indeed, this entire class of models is sometimes referred to as *tobit* models, a term coined by Goldberger (1964) to stand for "Tobin's probit." In the 1970s, a series of extensions of Tobin's original model appeared that stimulated a great deal of empirical and theoretical work. These include Grounau (1973), Heckman (1974, 1976), and Hausman and Wise (1977). See Amemiya (1985, Chapter 10) for an extensive review of this literature, and Breen (1996) for a good introduction.