16.3. Tobit Model

Truncation and censoring arise most often in econometrics in the linear regression model with normally distributed error, when only positive outcomes are completely observed. This model is called the Tobit model after Tobin (1958), who applied it to individual expenditures on consumer durable goods. The model in practice is usually too restrictive. It is nonetheless presented in some detail, as it provides the basis for more general models presented in subsequent sections of this chapter.

16.3.1. Tobit Model

The censored normal regression model, or **Tobit model**, is one with censoring from below at zero where the latent variable is linear in regressors with additive error that is normally distributed and homoskedastic. Thus

$$y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon, \tag{16.11}$$

where the error term

$$\varepsilon \sim \mathcal{N}[0, \sigma^2] \tag{16.12}$$

has variance σ^2 constant across observations. This implies that the latent variable $y^* \sim \mathcal{N}[\mathbf{x}'\boldsymbol{\beta}, \sigma^2]$. The observed y is defined by (16.2) with $L = 0$, so

$$y = \begin{cases} y^* & \text{if } y^* > 0, \\ - & \text{if } y^* \leq 0, \end{cases} \tag{16.13}$$

where – means that y is observed to be missing. No particular value of y is necessarily observed when $y^* \leq 0$, though in some settings such as durable goods expenditures we observe $y = 0$.

Equations (16.11) – (16.13) define the prototypical Tobit model analyzed by Tobin (1958). More generally, Tobit models begin with (16.11) and (16.12) for the latent variable but can have other censoring mechanisms including censoring from above, censoring from both below and above (the **two-limit Tobit model**), and interval-censored data. The results in this section are restricted to the censoring mechanism given in (16.13). The models of later sections are sometimes called generalized Tobit models.

The normalization $L = 0$ is not only natural in many settings, but some such normalization is necessary for a linear model with intercept and constant threshold parameter L. Then we observe y if $y^* > L$, or equivalently if $\beta_1 + \mathbf{x}_2'\boldsymbol{\beta}_2 + \varepsilon > L$ or $(\beta_1 - L) + \mathbf{x}_2'\boldsymbol{\beta}_2 + \varepsilon > 0$. Thus only the difference $(\beta_1 - L)$ is identified. More generally, the latent model $y^* = \mathbf{x}'\boldsymbol{\beta} + \varepsilon$ with variable censoring threshold $L = \mathbf{x}'\boldsymbol{\gamma}$ is observationally equivalent to the latent model $y^* = \mathbf{x}'(\boldsymbol{\beta} - \boldsymbol{\gamma}) + \varepsilon$ with fixed threshold $L = 0$. These results are a consequence of censoring arising in a linear model with additive error and do not carry over to nonlinear models, such as the preceding Poisson example.

Applying the general expression (16.7) for the censored density, here $f^*(y)$ is the $\mathcal{N}[\mathbf{x}'\boldsymbol{\beta}, \sigma^2]$ density and

$$\begin{aligned} F^*(0) &= \Pr[y^* \leq 0] \\ &= \Pr[\mathbf{x}'\boldsymbol{\beta} + \varepsilon \leq 0] \\ &= \Phi\left(-\mathbf{x}'\boldsymbol{\beta}/\sigma\right) \\ &= 1 - \Phi\left(\mathbf{x}'\boldsymbol{\beta}/\sigma\right), \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal cdf and the last equality uses symmetry of the standard normal distribution. Thus the censored density can be expressed as

$$f(y) = \left[\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y - \mathbf{x}'\boldsymbol{\beta})^2\right\}\right]^d \left[1 - \Phi\left(\frac{\mathbf{x}'\boldsymbol{\beta}}{\sigma}\right)\right]^{1-d}, \tag{16.14}$$

where the binary indicator d is defined in (16.6) with $L = 0$.

The Tobit MLE $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\sigma}^2)'$ maximizes the censored log-likelihood function (16.8). Given (16.14) this becomes

$$\begin{aligned} \ln \mathrm{L}_N(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{N} & \left\{ d_i \left(-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^2 \right) \right. \\ & \left. + (1 - d_i)\ln\left(1 - \Phi\left(\frac{\mathbf{x}_i'\boldsymbol{\beta}}{\sigma}\right)\right)\right\}, \end{aligned} \tag{16.15}$$

a mixture of discrete and continuous densities. The first-order conditions are

$$\begin{aligned} \frac{\partial \ln \mathrm{L}_N}{\partial \boldsymbol{\beta}} &= \sum_{i=1}^{N} \frac{1}{\sigma^2}\left(d_i(y_i - \mathbf{x}_i'\boldsymbol{\beta}) - (1 - d_i)\frac{\sigma\phi_i}{(1 - \Phi_i)} \right)\mathbf{x}_i = \mathbf{0} \\ \frac{\partial \ln \mathrm{L}_N}{\partial \sigma^2} &= \sum_{i=1}^{N} \left\{ d_i\left(-\frac{1}{2\sigma^2} + \frac{\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^2}{2\sigma^4} \right) + (1 - d_i)\frac{\phi_i \mathbf{x}_i'\boldsymbol{\beta}}{(1 - \Phi_i)}\frac{1}{2\sigma^3} \right\} = 0, \end{aligned} \tag{16.16}$$

using $\partial\Phi(z)/\partial z = \phi(z)$ where $\phi(\cdot)$ is the standard normal pdf, and with the definitions $\phi_i = \phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)$ and $\Phi_i = \Phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)$. As usual $\widehat{\boldsymbol{\theta}}$ is consistent if the density is correctly specified, that is, if the dgp is (16.11) and (16.12) and the censoring mechanism is (16.13). The MLE is asymptotic normal distributed with variance matrix given in, for example, Maddala (1983, p. 155) and Amemiya (1985, p. 373).

Tobin (1958) proposed ML estimation of the Tobit model and asserted that the usual ML theory applied. Amemiya (1973) provided a formal proof that the usual theory did apply, despite the mixed discrete–continuous nature of the censored density. The appendix of this classic paper of Amemiya details the asymptotic theory for extremum estimators presented in Section 5.3.

If data are truncated, rather than censored, from below at zero then the Tobit MLE $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\sigma}^2)'$ maximizes the truncated normal log-likelihood function

$$\ln \mathrm{L}_N(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{N} \left\{ -\frac{1}{2}\ln\sigma^2 - \frac{1}{2}\ln 2\pi - \frac{1}{2\sigma^2}\left(y_i - \mathbf{x}_i'\boldsymbol{\beta}\right)^2 - \ln\Phi\left(\mathbf{x}_i'\boldsymbol{\beta}/\sigma\right) \right\}, \tag{16.17}$$

obtained using (16.9) for y^* distributed as in (16.11) and (16.12).

16.3.2. Inconsistency of the Tobit MLE

A very major weakness of the Tobit MLE is its heavy reliance on distributional assumptions. If the error ε is either heteroskedastic or nonnormal the MLE is inconsistent.

This can be seen from the ML first-order conditions (16.16), which are a quite complicated function of variables including d_i, y_i, ϕ_i, and Φ_i. The first equation in (16.16) satisfies $\mathrm{E}[\partial \ln \mathrm{L}_N/\partial\boldsymbol{\beta}] = \mathbf{0}$, a necessary condition for consistency (see Section 5.3.7), if

$$\begin{aligned} \mathrm{E}[d_i] &= \Phi_i, \\ \mathrm{E}[d_i y_i] &= \Phi_i \mathbf{x}_i'\boldsymbol{\beta} + \sigma\phi_i. \end{aligned}$$

These moment conditions can be shown to hold if the dgp is (16.11) and (16.12) and the censoring mechanism is (16.13). However, they are unlikely to hold under any other specification of the dgp, as they rely heavily on both normality and homoskedasticity. For example, with *heteroskedastic errors* the estimator is inconsistent, since then $\mathrm{E}[d_i] = \Phi(\mathbf{x}_i'\boldsymbol{\beta}/\sigma_i) \neq \Phi_i$ unless $\sigma_i^2 = \sigma^2$.

Consistent estimation with heteroskedastic normal errors is possible by specifying a model for heteroskedasticity, say $\sigma_i^2 = \exp(\mathbf{z}_i'\boldsymbol{\gamma})$. For censoring from below at zero the log-likelihood $\ln \mathrm{L}_N(\boldsymbol{\beta}, \boldsymbol{\gamma})$ is that given in (16.15) with σ^2 replaced by $\exp(\mathbf{z}_i'\boldsymbol{\gamma})$. Consistency then requires normal errors and correct specification of the functional form of the heteroskedasticity.

Clearly, with censoring or truncation, distributional assumptions become important even for distributions somewhat robust to misspecification in the uncensored or untruncated case. Specification tests for the Tobit model are discussed in Section 16.3.7. In many censored data applications the Tobit model is not appropriate. More general models presented in subsequent sections of this chapter are instead used.

16.3.3. Censored and Truncated Means in Linear Regression

Censoring and truncation in the linear regression model (16.11) lead to observed dependent variable y that has distribution with conditional mean other than $\mathbf{x}'\boldsymbol{\beta}$, conditional variance other than σ^2 even if ε is homoskedastic, and distribution that is nonnormal even if ε is normally distributed. We present general results for linear regression in this section before specializing to normally distributed errors in Sections 16.3.4–

16.3.7. The results provide additional insights regarding the consequences of truncation and censoring and form the basis for non-ML estimation methods presented in later sections.

We begin with the truncated mean. The effects of truncation are intuitively predictable. Left-truncation excludes small values, so the mean should increase, whereas with right-truncation the mean should decrease. Since truncation reduces the range of variation, the variance should decrease.

For *left-truncation* at zero we only observe y if $y^* > 0$. If we suppress dependence of expectations on $\mathbf{x}$ for notational simplicity, the left-truncated mean becomes

$$\begin{aligned} \mathrm{E}[y] &= \mathrm{E}[y^*|y^* > 0] \\ &= \mathrm{E}\left[\mathbf{x}'\boldsymbol{\beta} + \varepsilon|\mathbf{x}'\boldsymbol{\beta} + \varepsilon > 0\right] \\ &= \mathrm{E}\left[\mathbf{x}'\boldsymbol{\beta}|\mathbf{x}'\boldsymbol{\beta} + \varepsilon > 0\right] + \mathrm{E}\left[\varepsilon|\mathbf{x}'\boldsymbol{\beta} + \varepsilon > 0\right] \\ &= \mathbf{x}'\boldsymbol{\beta} + \mathrm{E}\left[\varepsilon|\varepsilon > -\mathbf{x}'\boldsymbol{\beta}\right], \end{aligned} \tag{16.18}$$

where the second equality uses (16.11), and the last equality assumes ε is independent of $\mathbf{x}$. As expected the truncated mean exceeds $\mathbf{x}'\boldsymbol{\beta}$, since $\mathrm{E}[\varepsilon|\varepsilon > c]$ for any constant c will exceed $\mathrm{E}[\varepsilon]$.

For data *left-censored* at zero suppose we observe $y = 0$, rather than merely that $y^* \leq 0$. The censored mean is obtained by first conditioning the observable y on the binary indicator d defined in (16.6) with $L = 0$ and then unconditioning. Suppressing dependence on $\mathbf{x}$ for notational simplicity again, we have the left-censored mean

$$\begin{aligned} \mathrm{E}[y] &= \mathrm{E}_d[\mathrm{E}_{y|d}[y|d]] \\ &= \Pr[d=0] \times \mathrm{E}[y|d=0] + \Pr[d=1] \times \mathrm{E}[y|d=1] \\ &= 0 \times \Pr[y^* \leq 0] + \Pr[y^* > 0] \times \mathrm{E}[y^*|y^* > 0] \\ &= \Pr[y^* > 0] \times \mathrm{E}[y^*|y^* > 0], \end{aligned} \tag{16.19}$$

where $\Pr[y^* > 0] = 1 - \Pr[y^* \leq 0] = \Pr[\varepsilon > -\mathbf{x}'\boldsymbol{\beta}]$ is one minus the censoring probability and $\mathrm{E}[y^*|y^* > 0]$ is the truncated mean already derived in (16.18).

In summary, for the linear regression model with censoring or truncation from below at zero, the conditional means are given by

$$\begin{array}{ll} \text{latent variable:} & \mathrm{E}[y^*|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta} \\ \text{left-truncated (at 0):} & \mathrm{E}[y|\mathbf{x}, y > 0] = \mathbf{x}'\boldsymbol{\beta} + \mathrm{E}\left[\varepsilon|\varepsilon > -\mathbf{x}'\boldsymbol{\beta}\right], \\ \text{left-censored (at 0):} & \mathrm{E}[y|\mathbf{x}] = \Pr[\varepsilon > -\mathbf{x}'\boldsymbol{\beta}]\left\{\mathbf{x}'\boldsymbol{\beta} + \mathrm{E}\left[\varepsilon|\varepsilon > -\mathbf{x}'\boldsymbol{\beta}\right]\right\}. \end{array} \tag{16.20}$$

It is clear that even though the original conditional mean is linear, censoring or truncation leads to conditional means that are nonlinear so that OLS estimates will be inconsistent.

One possible approach to take is a parametric one of assuming a distribution for ε. This leads to expressions for $\mathrm{E}\left[\varepsilon|\varepsilon > -\mathbf{x}'\boldsymbol{\beta}\right]$ and $\Pr[\varepsilon > -\mathbf{x}'\boldsymbol{\beta}]$ and hence the truncated or censored conditional mean. We do this in the next section for normally distributed errors.

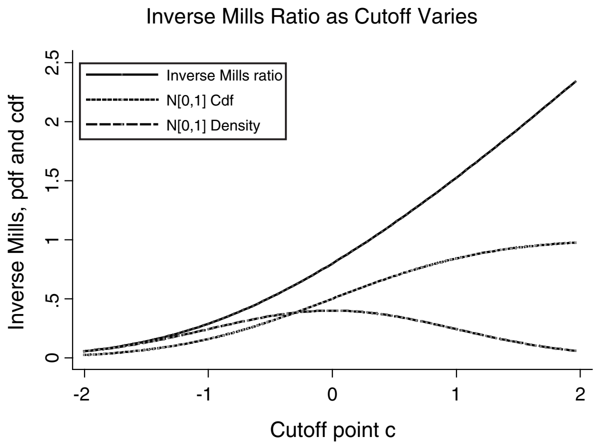

Figure 16.2: Inverse Mills ratio for the standard normal distribution as the censoring or cutoff point c increases. Standard normal cdf and density also plotted.

A second approach seeks to avoid or minimize such parametric assumptions. We consider this in a later section, but note here that regardless of the distribution for ε the truncated mean is a single-index model with correction term decreasing in $\mathbf{x}'\boldsymbol{\beta}$ since $\mathrm{E}\left[\varepsilon|\varepsilon > -\mathbf{x}'\boldsymbol{\beta}\right]$ is a monotonically decreasing function in $\mathbf{x}'\boldsymbol{\beta}$.

16.3.4. Censored and Truncated Means in the Tobit Model

For the Tobit model the regression error ε is normal and we use the following result, derived in Section 16.10.1.

Proposition 16.1 (Truncated Moments of the Standard Normal): *Suppose* $z \sim \mathcal{N}[0, 1]$. *Then the left-truncated moments of* z are

(i) $\mathrm{E}[z|z > c] = \phi(c)/[1 - \Phi(c)]$, *and* $\mathrm{E}[z|z > -c] = \phi(c)/\Phi(c)$,

(ii) $\mathrm{E}[z^2|z > c] = 1 + c\phi(c)/[1 - \Phi(c)]$, *and*

(iii) $\mathrm{V}[z|z > c] = 1 + c\phi(c)/[1 - \Phi(c)] - \phi(c)^2/[1 - \Phi(c)]^2$

Result (i) of Proposition 16.1 is shown in Figure 16.2. We consider truncation of $z \sim \mathcal{N}[0, 1]$ from below at c, where c ranges from -2 to 2. The lowest curve is the standard normal density $\phi(c)$ evaluated at c. The middle curve is the standard normal cdf $\Phi(c)$ evaluated at c and gives the probability of truncation when truncation is at c. This probability is approximately 0.023 at $c = -2$ and 0.977 at $c = 2$. The upper curve gives the truncated mean $\mathrm{E}[z|z > c] = \phi(c)/[1 - \Phi(c)]$. As expected this is close to $\mathrm{E}[z] = 0$ for $c = -2$, since then there is little truncation, and $\mathrm{E}[z|z > c] > c$. What is not expected a priori is that $\phi(c)/[1 - \Phi(c)]$ is approximately linear, especially for $c > 0$. Moments when truncation is from above can be obtained using, for example, $\mathrm{E}[z|z < c] = -\mathrm{E}[-z| - z > -c] = -\phi(c)/\Phi(c)$.

Appling this result to (16.18), the error term has truncated mean

$$\begin{aligned} \mathrm{E}\left[\varepsilon|\varepsilon > -\mathbf{x}'\boldsymbol{\beta}\right] &= \sigma \mathrm{E}\left[\tfrac{\varepsilon}{\sigma}|\tfrac{\varepsilon}{\sigma} > \tfrac{-\mathbf{x}'\boldsymbol{\beta}}{\sigma}\right] \\ &= \sigma\phi(-\tfrac{\mathbf{x}'\boldsymbol{\beta}}{\sigma})/[1-\Phi(-\tfrac{\mathbf{x}'\boldsymbol{\beta}}{\sigma})] \\ &= \sigma\phi(\tfrac{\mathbf{x}'\boldsymbol{\beta}}{\sigma})/[\Phi(\tfrac{\mathbf{x}'\boldsymbol{\beta}}{\sigma})] \\ &= \sigma\lambda(\tfrac{\mathbf{x}'\boldsymbol{\beta}}{\sigma}), \end{aligned} \tag{16.21}$$

where the second line uses Proposition 16.1, the third line uses symmetry about zero of $\phi(z)$, and we define

$$\lambda(z) = \frac{\phi(z)}{\Phi(z)}. \tag{16.22}$$

We follow the definition and terminology of Amemiya (1985) and many others in defining $\lambda(\cdot)$ as in (16.22) and calling it the **inverse Mills ratio**. From Johnson and Kotz (1970, p. 278), Mills actually tabulated the ratio $(1-\Phi(z))/\phi(z)$ whose inverse $\phi(z)/[1-\Phi(z)] = \phi(z)/\Phi(-z)$ is the hazard function of the normal distribution. Some authors therefore instead write (16.21) as $\mathrm{E}\left[\varepsilon|\varepsilon > -\mathbf{x}'\boldsymbol{\beta}\right] = \sigma\lambda^*(-\mathbf{x}'\boldsymbol{\beta}/\sigma)$, where $\lambda^*(z) = \phi(z)/\Phi(-z)$ is referred to as the inverse Mills ratio.

Also, $\Pr[\varepsilon > -\mathbf{x}'\boldsymbol{\beta}] = \Pr[-\varepsilon < \mathbf{x}'\boldsymbol{\beta}] = \Pr[-\varepsilon/\sigma < \mathbf{x}'\boldsymbol{\beta}/\sigma] = \Phi(\mathbf{x}'\boldsymbol{\beta}/\sigma)$. Then the conditional means in (16.20) specialize to

$$\begin{array}{ll} \text{latent variable:} & \mathrm{E}[y^*|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta}, \\ \text{left-truncated (at 0):} & \mathrm{E}[y|\mathbf{x}, y>0] = \mathbf{x}'\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}'\boldsymbol{\beta}/\sigma), \\ \text{left-censored (at 0):} & \mathrm{E}[y|\mathbf{x}] = \Phi(\mathbf{x}'\boldsymbol{\beta}/\sigma)\mathbf{x}'\boldsymbol{\beta} + \sigma\phi(\mathbf{x}'\boldsymbol{\beta}/\sigma). \end{array} \tag{16.23}$$

The variance is similarly obtained (see Exercise 16.1). Defining $w = \mathbf{x}'\boldsymbol{\beta}/\sigma$, we have

$$\begin{array}{ll} \text{latent variable:} & \mathrm{V}[y^*|\mathbf{x}] = \sigma^2, \\ \text{left-truncated (at 0):} & \mathrm{V}[y|\mathbf{x}, y>0] = \sigma^2\left[1 - w\lambda(w) - \lambda(w)^2\right], \\ \text{left-censored (at 0):} & \mathrm{V}[y|\mathbf{x}] = \sigma^2\Phi(w)\left\{w^2 + w\lambda(w) + 1 - \Phi(w)[w+\lambda(w)]\right\}^2. \end{array} \tag{16.24}$$

Clearly truncation and censoring induce heteroskedasticity, and for truncation $\mathrm{V}[y|\mathbf{x}] < \sigma^2$ so that truncation reduces variability, as expected.

These results assume normal errors. Maddala (1983, p. 369) gives results similar to Proposition 16.1 for the log-normal, logistic, uniform, Laplace, exponential, and gamma distributions.

16.3.5. Marginal Effects in the Tobit Model

The marginal effect is the effect on the conditional mean of the dependent variable of changes in the regressors. This effect varies according to whether interest lies in the latent variable mean $\mathbf{x}'\boldsymbol{\beta}$ or the truncated or censored means given in (16.23).

Differentiating each with respect to $\mathbf{x}$ yields

$$\begin{aligned}
&\text{latent variable:} && \partial \mathrm{E}[y^*|\mathbf{x}]/\partial \mathbf{x} = \boldsymbol{\beta}, \\
&\text{left-truncated (at 0):} && \partial \mathrm{E}[y, y > 0|\mathbf{x}]/\partial \mathbf{x} = \{1 - w\lambda(w) - \lambda(w)^2\}\boldsymbol{\beta}, \\
&\text{left-censored (at 0):} && \partial \mathrm{E}[y|\mathbf{x}]/\partial \mathbf{x} = \Phi(w)\boldsymbol{\beta},
\end{aligned} \tag{16.25}$$

where $w = \mathbf{x}'\boldsymbol{\beta}/\sigma$ and we use $\partial\Phi(z)/\partial z = \phi(z)$ and $\partial\phi(z)/\partial z = -z\phi(z)$. The simple expression for the censored mean is obtained after some manipulation. It can be decomposed into two effects, one for $y = 0$ and one for $y > 0$ (see McDonald and Moffitt, 1980).

In some cases truncation or censoring is just an artifact of data collection, so the truncated and censored means are of no intrinsic interest and we are interested in $\partial \mathrm{E}[y^*|\mathbf{x}]/\partial \mathbf{x} = \boldsymbol{\beta}$. For example, with top-coded earnings data we are clearly interested in measuring the effect of schooling on mean earnings rather than earnings of those not top-coded.

In other cases truncation or censoring has behavioral implications. In a model for hours worked, for example, the three marginal effects in (16.25) correspond to the effect of a change in a regressor on, respectively, (1) desired hours of work, (2) actual hours of work for workers, and (3) actual hours of work for workers and nonworkers. For (1) we clearly need an estimate of $\boldsymbol{\beta}$, but for (2) and (3) OLS slope coefficients, although inconsistent for $\boldsymbol{\beta}$, may actually provide a reasonable crude estimate of the marginal effect since the truncated and censored means are still fairly linear in $\mathbf{x}$.

16.3.6. Alternative Estimators for the Tobit Model

In addition to the MLE, consistent estimation is possible by NLS based on the correct expression for the truncated or censored mean. We consider the NLS estimator and other least-squares estimators.

NLS Estimator

The results in (16.23) can be used to permit consistent estimation of the Tobit model parameters by NLS. For example, with truncated data we minimize

$$S_N(\boldsymbol{\beta}, \sigma^2) = \sum_{i=1}^{N} \left(y_i - \mathbf{x}_i'\boldsymbol{\beta} - \sigma\lambda(\mathbf{x}_i'\boldsymbol{\beta}/\sigma)\right)^2$$

with respect to both $\boldsymbol{\beta}$ and σ^2, but then perform inference controlling for the heteroskedasticity given in (16.24). A similar estimator can be obtained for censored data.

This estimator is not used in practice. Consistency requires correct specification of the truncated mean, which from (16.21) requires both normality and homoskedasticity of the errors. One might as well estimate by ML since this relies on assumptions just as strong and is fully efficient. Moreover, in practice the NLS estimator can be imprecise. From Figure 16.2 it is clear that $\lambda(\mathbf{x}'\boldsymbol{\beta}/\sigma)$ is approximately linear in $\mathbf{x}'\boldsymbol{\beta}/\sigma$, leading to near collinearity because $\mathbf{x}$ is also a regressor. In Section 16.5 we consider models that permit correction terms similar to $\sigma\lambda(\mathbf{x}'\boldsymbol{\beta}/\sigma)$ in (16.23) that have the advantage of depending in part on regressors other than those in $\mathbf{x}$.

Heckman Two-Step Estimator

From (16.23) the truncated (at zero) mean is

$$\mathrm{E}[y|\mathbf{x}] = \mathbf{x}'\boldsymbol{\beta} + \sigma\lambda(\mathbf{x}'\boldsymbol{\beta}/\sigma). \tag{16.26}$$

Rather than use NLS, this can be estimated in the following two-step procedure if censored data are available. First, for the full sample do probit regression of d on $\mathbf{x}$, where the binary variable d equals one if $y > 0$ is observed, to give consistent estimate $\widehat{\boldsymbol{\alpha}}$, where $\boldsymbol{\alpha} = \boldsymbol{\beta}/\sigma$. Second, for the truncated sample do OLS regression of y on $\mathbf{x}$ and $\lambda(\mathbf{x}'\widehat{\boldsymbol{\alpha}})$ to give consistent estimates of $\boldsymbol{\beta}$ and σ.

This estimation procedure, due to Heckman (1976, 1979), is presented in Section 16.5.4 where it is applied to the more general sample selection model. Section 16.10.2 derives the standard error of $\widehat{\boldsymbol{\beta}}$ that accounts for the regressor $\lambda(\mathbf{x}'\widehat{\boldsymbol{\alpha}})$ depending on estimated parameters and for heteroskedasticity induced by truncation.

OLS Estimation of the Tobit Model

The OLS estimates using censored or truncated data are inconsistent for $\boldsymbol{\beta}$. This is because the censored and truncated means given in (16.23) are not equal to $\mathbf{x}'\boldsymbol{\beta}$, violating the essential condition for consistency of OLS.

For censored data, OLS provides a linear approximation to the nonlinear censored regression curve. It is clear from Figure 16.1 and (16.25) that this line is flatter than the regression line for uncensored data, which has slope equal to the true slope parameter. Goldberger (1981) showed analytically that if y and $\mathbf{x}$ are joint normally distributed and there is censoring from below at zero, then the OLS slope parameters converge to p times the true slope parameter, where p is the fraction of the sample with positive values of y. These conditions are restrictive but were relaxed somewhat by Ruud (1986). In practice this proportionality result provides a good empirical approximation to the inconsistency of OLS if a Tobit model is instead appropriate.

Similarly, with truncation the regression line is flatter than the untruncated regression line. Goldberger (1981) obtained an analytical result similar to that for the censored case. If y and $\mathbf{x}$ are joint normally distributed and there is censoring from below at zero, then the OLS slope parameters converge to a multiple of the true slope parameter. The multiple, the expression for which is quite lengthy, lies between zero and one, and the shrinkage is the same for all slope coefficients. Truncated OLS therefore understates the absolute magnitude of the true slope parameters.

16.3.7. Specification Tests for the Tobit Model

Given the fragility of the Tobit model it is good practice to test for distributional misspecification. There are four broad strategies.

The first approach is to nest the Tobit model within a richer parametric model and apply a Wald, LR, or LM test. Since the null hypothesis model, the Tobit model, is most easily estimated it is natural to use LM tests. This is particularly straightforward for testing against heteroskedasticity of the form $\sigma_i^2 = \exp(\mathbf{x}_i'\boldsymbol{\alpha})$ in the censored