Truncated Regression Models

It is quite simple to estimate a truncated regression model by maximum likelihood if the distribution of the error terms in the latent variable model is assumed to be known. By far the most common assumption is that the error terms are normally, independently, and identically distributed, as in (11.66). We restrict our attention to this special case.

If the regression function for the latent variable model is $\boldsymbol{X}_t\boldsymbol{\beta}$, the probability that y_t° is included in the sample is

$$\begin{aligned}\Pr(y_t^\circ \geq 0) &= \Pr(\boldsymbol{X}_t\boldsymbol{\beta} + u_t \geq 0)\\ &= 1 - \Pr(u_t < -\boldsymbol{X}_t\boldsymbol{\beta}) = 1 - \Pr(u_t/\sigma < -\boldsymbol{X}_t\boldsymbol{\beta}/\sigma)\\ &= 1 - \Phi(-\boldsymbol{X}_t\boldsymbol{\beta}/\sigma) = \Phi(\boldsymbol{X}_t\boldsymbol{\beta}/\sigma).\end{aligned}$$

When $y_t^\circ \geq 0$ and y_t is observed, the density of y_t is proportional to the density of y_t°. Otherwise, the density of y_t is 0. The factor of proportionality, which is needed to ensure that the density of y_t integrates to unity, is the inverse of the probability that $y_t^\circ \geq 0$. Therefore, the density of y_t can be written as

$$\frac{\sigma^{-1}\phi\big((y_t - \boldsymbol{X}_t\boldsymbol{\beta})/\sigma\big)}{\Phi(\boldsymbol{X}_t\boldsymbol{\beta}/\sigma)}.$$

This implies that the loglikelihood function, which is the sum over all t of the log of the density of y_t conditional on $y_t^\circ \geq 0$, is

$$\begin{aligned}\ell(\boldsymbol{y}, \boldsymbol{\beta}, \sigma) = -\frac{n}{2}\log(2\pi) - n\log(\sigma) - \frac{1}{2\sigma^2}\sum_{t=1}^{n}(y_t - \boldsymbol{X}_t\boldsymbol{\beta})^2\\ - \sum_{t=1}^{n}\log\Phi(\boldsymbol{X}_t\boldsymbol{\beta}/\sigma).\end{aligned} \tag{11.67}$$

Maximization of expression (11.67) is generally not difficult. Even though the loglikelihood function is not globally concave, there is a unique MLE; see Orme and Ruud (2002).

The first three terms in expression (11.67) comprise the loglikelihood function that corresponds to OLS regression; see equation (10.10). The last term is minus the summation over all t of the logarithms of the probabilities that an observation with regression function $\boldsymbol{X}_t\boldsymbol{\beta}$ belongs to the sample. Since these probabilities must be less than 1, this term must always be positive. It can be made larger by making the probabilities smaller. Thus the maximization algorithm chooses the parameters in such a way that these probabilities are smaller than they would be for the OLS estimates. The presence of this fourth term therefore causes the ML estimates of $\boldsymbol{\beta}$ and σ to differ, often substantially, from their least-squares counterparts, and it ensures that the ML estimates are consistent.