10 Censored Models

10.1 Censoring

Suppose that y_i^* is a latent variable, and the observed variable is censored

$$\begin{aligned} y_i &= y_i^* 1\left(y_i^* > 0\right)) \\ &= \begin{cases} 0 & y_i^* \leq 0 \\ y_i^* & y_i^* > 0 \end{cases} \end{aligned}$$

Notationally we have set the censoring point at zero, but this is not essential.

If the distribution of y_i^* is nonparametric, moments of y_i^* are unidentified. We don't observe the full range of support for y_i^*, so anything can happen in that part of the distribution. It follows that moment restrictions are inherently unidentified. When there is censoring, we should be very cautious about moment restriction models.

In contrast, (some) quantiles are identified. Let $q_a(y_i^*)$ and $q_a(y_i)$ denote the α'th quantiles of y_i^* and y_i.

If $P(y_i^* \leq 0) < \alpha$, then $q_\alpha(y_i^*) = q_\alpha(y_i)$. That is, so long as there is less than α percent censored, censoring does not affect quantiles.

Furthermore, if If $P(y_i^* \leq 0) \geq \alpha$ then $q_\alpha(y_i) = 0$. (e.g., if there is 30% censoring, then then quantiles below 30% are identically zero).

This means that we have the relationship:

$$q_\alpha(y_i) = \max(q_\alpha(y_i^*\ ), 0)$$

It follows that we can consistently (and efficiently) estimate quantiles above the α'th on the observed data y_i. These observations lead to the strong conclusion that in the presence of censoring, we should identify parameters through quantile restrictions, not moment restrictions.

Of particular interest is the median $Med\left(y_i^*\right)$. We have

$$Med\left(y_i\right) = \max(Med\left(y_i^*\right), 0)$$

10.2 Powell's CLAD Estimator

The Tobit or censored regression model is

$$\begin{aligned} y_i^* &= X_i'\beta + e_i \\ y_i &= y_i^* 1\left(y_i^* > 0\right)) \end{aligned}$$

The classic Tobit estimator for β is MLE when e_i is independent of X_i and $N(0, \sigma^2)$.

Powell (1984, Journal of Econometrics) made the brilliant observation that when e_i is nonparametric, β is not identified through moment restrictions.

Instead, identify $X_i'\beta$ as the conditional median of y_i, so

$$Med\left(y_i^* \mid X_i\right) = X_i'\beta$$

As we showed in the previous section

$$\begin{aligned} Med\left(y_i \mid X_i\right) &= \max(Med\left(y_i^* \mid X_i\right), 0) \\ &= \max(X_i'\beta, 0) \end{aligned}$$

Thus the conditional median is a specific nonlinear function of the single index $\beta' X_i$.

This shows that the censored observation obeys the nonlinear median regression model

$$\begin{aligned} y_i &= \max(X_i'\beta, 0) + \varepsilon_i \\ Med\left(\varepsilon_i \mid X_i\right) &= 0 \end{aligned}$$

We know that the appropriate method to estimate conditional medians is by least absolute deviations (LAD). This applies as well to nonlinear models. Hence Powell suggested the criterion

$$S_n(\beta) = \sum_{i=1}^{n} \left|y_i - \max(X_i'\beta, 0)\right|$$

or equivalently we can use the criterion

$$S_n(\beta) = \sum_{i=1}^{n} 1\left(X_i'\beta > 0\right)\left|y_i - X_i'\beta\right|.$$

The estimator $\hat{\beta}$ which minimizes $S_n(\beta)$ is called the censored least absolute deviations (CLAD) estimator. The estimator satisfies the asymptotic FOC

$$\sum_{i=1}^{n} 1\left(X_i'\hat{\beta} > 0\right) x_i \operatorname{sgn}\left(y_i - X_i'\hat{\beta}\right) = 0$$

This is the same as the FOC for LAD, but only for the observations for which $X_i'\hat{\beta} > 0$.

Minimization of $S_n(\beta)$ is somewhat more tricky than standard LAD. Bushinsky (PhD dissertation) worked out numerical methods to solve this problem

Powell showed that it has the asymptotic distribution

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \to_d N(0, V)$$

$$
\begin{aligned}
V &= Q^{-1}\Omega Q^{-1} \\
\Omega &= \mathrm{E}\left(1_i X_i X_i'\right) \\
Q &= 2\mathrm{E}\left(f\left(0 \mid X_i\right) 1_i X_i X_i'\right) \\
1_i &= 1\left(X_i'\beta > 0\right)
\end{aligned}
$$

where $f(0 \mid x)$ is the conditional density of e_i given $X_i = x$ at the origin.

The derivation of this result is not much different from that for standard LAD regression. Since the criterion function is not smooth with respect to β, you need to use an empirical process approach, as outlined for example in section 7 of Newey and McFadden's Handbook chapter.

Identification requires that Ω and Q are full rank. This requires that there is not "too much" censoring. As the censoring rate increases, the information in Ω diminishes and precision falls.

10.3 Variance Estimation

When e_i is independent of X_i, then $f(0 \mid x) = f(0)$ and $V = \left(4f\left(0\right)^2 \Omega\right)^{-1}$. Practical standard error estimation seems to focus on estimating V under this assumption.

$$
\begin{aligned}
\hat{V} &= \left(4\hat{f}\left(0\right)^2 \hat{\Omega}\right)^{-1} \\
\hat{\Omega} &= \frac{1}{n}\sum_{i=1}^{n} 1\left(X_i'\hat{\beta} > 0\right) X_i X_i'
\end{aligned}
$$

The difficult part is $f(0)$, in part because $\hat{e}_i$ is only observed for observations with $y_i > 0$.

Hall and Horowitz (1990, Econometric Theory) recommend

$$
\hat{f}(0) = \frac{\sum_{i=1}^{n} k\left(\dfrac{\hat{e}_i}{h}\right) 1\left(y_i > 0\right)}{h\sum_{i=1}^{n} G\left(\dfrac{X_i'\hat{\beta}}{h}\right)}
$$

where $\hat{e}_i = y_i - X_i'\hat{\beta}$, h is a bandwidth, $k(u)$ is a symmetric kernel, and $G(u)$ is integrated kernel. They find that the optimal rate is $h \sim n^{-1/5}$ if k is a second-order kernel, and $h \sim n^{-1/(2\nu+1)}$ if k is a ν'th order kernel. Their paper has an expression for the optimal bandwidth, and discuss possible methods to estimate the bandwidth, but do not present a fully automatic bandwidth method.

An obvious alternative to asymptotic methods is the bootstrap. It is quite common to use bootstrap percentile methods to compute standards errors, confidence intervals, and p-values for LAD, quantile estimation, and CLAD estimation.

10.4 Khan and Powell's Two-Step Estimator

$$S_n(\beta) = \sum_{i=1}^{n} 1\left(X_i'\beta > 0\right)\left|y_i - X_i'\beta\right|.$$

In this criterion, the coefficient β plays two roles. Khan and Powell (2001, Journal of Econometrics) suggest this double role induces bias in finite samples, and this can be avoided by a two-step estimator.

They suggest first estimating $\tilde{\beta}$ using a semiparametric binary choice estimator, and using this to define the observations for trimming. The second-stage criterion is then

$$S_n(\beta) = \sum_{i=1}^{n} 1\left(X_i'\tilde{\beta} > 0\right)\left|y_i - X_i'\beta\right|.$$

(In the theoretical treatment, the indictor function is replaced with a smooth weighting function, but they claim this is only to make the theory easy, and they use the indicator function in their simulations.) The second-stage estimator minimizes this criterion, which is just LAD on the trimmed sub-sample. Khan and Powell argue that this two-step estimator falls in the class of Andrews' MINPIN estimators, so the asymptotic distribution is identical to Powell's estimator.

10.5 Newey and Powell's Weighted CLAD Estimator

When e_i is not independent of X_i, the asymptotic covariance matrix of the CLAD estimator suggests that it is inefficient and can be improved. Newey and Powell (1990, Econometric Theory) compute the semiparametric efficiency bound, and find that it is attained by the estimator minimizing the weighted criterion

$$\begin{aligned} S_n(\beta) &= \sum_{i=1}^{n} w_i \left|y_i - \max(X_i'\beta, 0)\right| \\ w_i &= 2f\left(0 \mid X_i\right) \end{aligned}$$

The estimator $\hat{\beta}$ which minimizes this criterion is a weighted CLAD estimator, and the authors show that it has the asymptotic distribution

$$\sqrt{n}\left(\hat{\beta} - \beta\right) \to_d N(0, V)$$

$$V = \left(4\mathrm{E}\left(f\left(0 \mid X_i\right)^2 1_i X_i X_i'\right)\right)^{-1}$$

The conditional density plays a similar role to the conditional variance for GLS regression.

This efficiency result is general to median regression, not just censored regression. That is, the weighted LAD estimator achieves the asymptotic efficiency bound for median regression. The unweighted estimator is efficient when $f(0 \mid x) = f(0)$ (essentially, when e_i is independent of X_i).

Feasible versions of this estimator are challenging to construct. Newey and Powell suggest a method based on nearerst neighbor regression estimation of the conditional distribution function.

I don't know if this has been noticed elsewhere, but here is a useful observation.

Suppose that the error e_i only depends on x_i through a scale effect. That is

$$e_i = \sigma(X_i) z_i$$

where z_i is independent of X_i, with density $f_z(z)$ and median zero. Then the conditional density of e_i given $X_i = x$ is

$$f(e \mid x) = \frac{1}{\sigma(x)} f_z\left(\frac{e}{\sigma(x)}\right)$$

so at the origin

$$f(0 \mid x) = \frac{1}{\sigma(x)} f_z(0)$$

Thus the optimal weighting is $w_i \sim \sigma(X_i)^{-1}$, which takes the same form as in the case of GLS in regression. The interpretation of $\sigma(x)$ is a bit different (it is identified on median restrictions).

10.6 Nonparametric Censored Regression

The models discussed in the previous sections assume that the conditional median is linear in X_i – a highly parametric assumption. It would be desireable to extend the censored regression model to allow for nonparametric median functions. A nonparametric model would take the form

$$\begin{aligned} Med(y_i^* \mid X_i) &= g(X_i) \\ y_i &= y_i^* 1(y_i^* > 0)) \end{aligned}$$

with g nonparametric. The conditional median for the observed dependent variable is

$$Med(y_i \mid X_i) = \max(g(X_i), 0).$$

We can define the conditional median function

$$g^*(x) = \max(g(x), 0).$$

Since $g(x)$ is nonparametric then so is $g^*(x)$, although it does satisfy $g^*(x) \geq 0$.

A feasible approach to estimate $g^*(x)$ is to simply use standard nonparametric median regression. It is unclear if any information is lost in ignoring the censoring. My only thought is that function $g^*(x)$ will typically have a "kink" at $g(x) = 0$, and this is smoothed over by nonparametric methods, which suggests inefficiency.

An alternative suggestion is Lewbel and Linton (Econometrica, 2002). They impose the strong assumption that the error $e_i = y_i^* - g(X_i)$ is independent of X_i, and develop nonparametric estimates of $g(x)$ using kernel methods.