Censored Regression Model

James Heckman
University of Chicago

Econ 312
This Draft, April 22, 2007
Outcome Equation:

\[ Y = X\beta + U \quad E(U) = 0 \]

Choice Equation:

\[ D^* = Z\gamma - V \quad \text{Index function} \]

\[ \text{Var}(V) = \sigma^2_V \]

\[ D = 1(D^* \geq 0) \]
Assume \((X, Z) \perp \perp (U, V)\).

\[
\Pr(D = 1 \mid Z = z) = \Pr(z\gamma > V) \\
= \Pr \left( \frac{z\gamma}{\sigma_V} > \frac{V}{\sigma_V} \right)
\]

Let

\[F_V(t) = \Pr(V \leq t).\]

Then

\[
\Pr(D = 1 \mid Z = z) = F_V \left( \frac{z\gamma}{\sigma_V} \right).
\]
Censoring

- Observe $Y$ if $D = 1$.

$$E(Y \mid D = 1, X = x, Z = z) = X\beta + E(U \mid D = 1, X, Z)$$

- From the index model, we obtain

$$E(Y \mid D = 1, X = x, Z = z) = X\beta + E\left( U \left| \frac{z\gamma}{\sigma_V} > \frac{V}{\sigma_V}, X = x, Z = z \right. \right)$$
Let $V/\sigma_V = V^*$.

Look at the control function.

The joint density of $(U, V/\sigma_V)$ is

$$f_{U,V^*}(t_1, t_2)$$
From (1), this density does not depend on \( X = x, Z = z \).

\[
E(U \mid z\gamma > V^*, X = x, Z = z) = \frac{\int u \int_{-\infty}^{z\gamma/\sigma_V} f_{U,V^*}(u, v^*) \, du \, dv^*}{\int_{-\infty}^{z\gamma/\sigma_V} f_{V^*}(v^*) \, dv^*}.
\]

Now

\[P(z) = \Pr(D = 1 \mid Z = z) = F_{v^*} \left( \frac{z\gamma}{\sigma_V} \right).\]
Assume $V^*$ is absolutely continuous:

$$\frac{Z\gamma}{\sigma_V} = F_{V^*}^{-1}(P(z))$$

$$E(U \mid Z\gamma > V^*, X = x, Z = z)$$

$$= \frac{1}{P(z)} \int u \int_{-\infty}^{\frac{z\gamma}{\sigma_V}} f_{U,V^*}(u, v^*) \, dv^* \, du$$

$$= \frac{1}{P(z)} \int u \int_{-\infty}^{F_{V^*}^{-1}(P(z))} f_{U,V^*}(u, v^*) \, dv^* \, du$$

$$= K(P(z))$$

where $K$ depends on the parameters of $f_{U,V^*}$ but not $X, Z$. 
Now as $P(z) \to 1$, $K(P(z)) \to 0$.

For the normal case, as $P(z) \to 0$, $K(P(z)) \to \infty$.

Normal case

$$(U, V) \sim N(0, \Sigma)$$

$$E(Y \mid X = x, Z = z, D = 1) = X\beta + \text{Cov}(U, V^*) \lambda \left( \frac{z\gamma}{\sigma_V} \right)$$

$$= X\beta + \text{Cov}(U, V^*) \lambda \left( \Phi^{-1}P(z) \right)$$