

Truncation and Censoring: Lecture XXXV

Charles B. Moss ¹

¹University of Florida

April 11, 2012

- 1 Readings
- 2 Tale of Two Tails
 - Production Frontier
 - Consumer Behavior
 - Probability Density Functions
- 3 Truncated Regressions
 - Truncated Normal
 - Truncated Normal Regression
 - Stochastic Frontier Model
- 4 Censored Regression

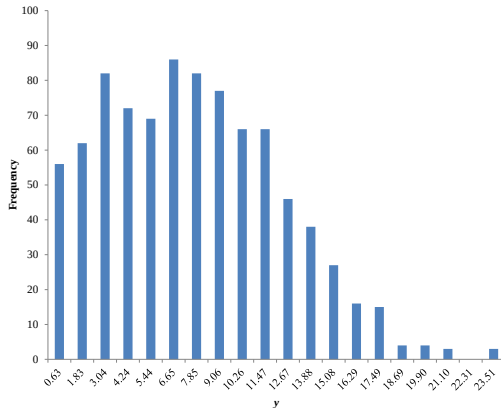
Readings

- * Cameron, A. C. and P. K. Trivedi. 2010. *Microeconometrics Using Stata Revised Edition*. Stata Press. (Chapter 16).
- * Greene, W.H. 2012. *Econometric Analysis*. Seventh Edition (Chapter 19 [Section 19.1 - 19.3 pp. 833 – 860])

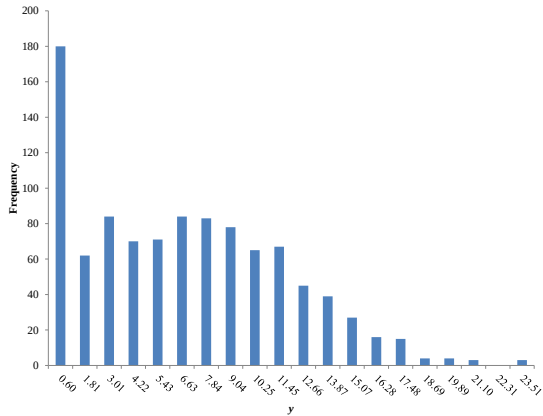
Tale of Two Tails

- Consider two different histograms of a random variable.
 - 1 The second example is the variation in production – production frontier minus actual production level.
 - 2 The first example is the result of consumer behavior – The consumer chooses to buy some quantity of a good (x_{1t}) based on prices and income.

Production Frontier



Consumer Behavior



Probability Density Functions

- The production frontier does not have a "bump" at the lower bound, the probability density function is smooth

$$\begin{aligned} f(y|\beta, \sigma^2) &= \tilde{f}(y|\beta, \sigma^2) \quad \exists: y \geq 0 \\ f(y|\beta, \sigma^2) &= 0 \quad \exists: y < 0 \end{aligned} \quad (1)$$

- The consumer behavior model has an irregular "bump" (actually a non-zero probability value) at the zero

$$\begin{aligned} f(y|\beta, \sigma^2) &= K \quad \exists: y = 0 \\ f(y|\beta, \sigma^2) &= \tilde{f}(y|\beta, \sigma^2) \quad \exists: y > 0 \\ f(y|\beta, \sigma^2) &= 0 \quad \exists: y < 0 \end{aligned} \quad (2)$$

Truncated Regressions

- The production frontier is the result of a **Truncated Distribution**:

Theorem 19.1: Density of a Truncated Random Variable *If a continuous random variable x has a pdf $f(x)$ and a is a constant, then*

$$f(x|x > a) = \frac{f(x)}{\text{Prob}(x > a)} \quad (3)$$

The proof follows from the definition of conditional probability and amounts merely to scaling the density so that it integrates to one over the range above a . Note that the truncated distribution is a conditional distribution (Greene p. 834).

Truncated Normal

- Assuming that y is normally distributed with a mean of μ and a standard deviation of σ the $\text{Prob}(y > a)$ becomes

$$\text{Prob}(y > a) = 1 - \Phi\left(\frac{a - \mu}{\sigma}\right) = 1 - \Phi(\alpha) \quad (4)$$

$$\alpha = \frac{a - \mu}{\sigma}$$

and $\Phi(\cdot)$ is the standard normal cumulative distribution function.

Truncated Normal – Continued

- The truncated distribution function is then

$$\begin{aligned}
 f(y|y > a) &= \frac{f(y)}{1 - \Phi(\alpha)} \\
 &= \frac{(2\pi)^{-1/2} \exp \left[- (y - \mu)^2 / (2\sigma^2) \right]}{1 - \Phi(\alpha)} \\
 &= \frac{\frac{1}{\sigma} \phi \left(\frac{y - \mu}{\sigma} \right)}{1 - \Phi(\alpha)}
 \end{aligned} \tag{5}$$

Truncated Normal Regression

- Expanding on this formulation

$$\begin{aligned}\mu(x) &= x_i\beta \\ y_i &= x_i\beta + \epsilon_i\end{aligned}\tag{6}$$

- The truncation of the normal changes the expectation operator slightly

$$\begin{aligned}\mathbb{E}[y|y > a] &= \mu + \sigma\lambda(\alpha) \\ \lambda(\alpha) &= \frac{\phi(\alpha)}{1 - \Phi(\alpha)}\end{aligned}\tag{7}$$

Stochastic Frontier Model

- The truncated regression model is sometimes used to estimate the technical inefficiency of production. In the Cobb-Douglas form

$$\ln(y_i) = \beta_0 + \sum_{k=1}^K \beta_k \ln(x_{ki}) - \epsilon_i \quad (8)$$

where ϵ_t is "one-sided" or more specifically non-negative.

- Conceptually, we add a "two-sided" error term (ν_i) which accounts for things like excluded variables

$$\ln(y_i) = \beta_0 + \sum_{k=1}^K \beta_k \ln(x_{ki}) - \epsilon_i + \nu_i \quad (9)$$

Stochastic Frontier Model – Continued

- Folding these two concepts together

$$\ln(h(\zeta_i | \beta, \lambda, \sigma)) = -\ln(\sigma) + \frac{1}{2} \ln\left(\frac{2}{\pi}\right) - \frac{1}{2} \left(\frac{\zeta_i}{\sigma}\right)^2 + \ln\left(\Phi\left(\frac{-\zeta_i \lambda}{\sigma}\right)\right)$$

$$\zeta_i = \nu_i - \epsilon_i, \lambda = \frac{\sigma_\epsilon}{\sigma_\nu}, \sigma = \left(\sigma_\nu^2 + \sigma_\epsilon^2\right)^{1/2}$$
(10)

- We will develop this model more fully in AEB 6184 – Lecture 11 (<http://ricardo.ifas.ufl.edu/aeb6184.production/Lecture11-2011.pdf>).

Censored Regression

- The **Censored Regression** accounts for the positive mass at zero in the Consumer Behavior model

$$\begin{aligned} y_i^* &= x_i\beta + \epsilon_i \\ y_i &= 0 \text{ if } y_i^* \leq 0 \\ y_i &= y_i^* \text{ if } y_i^* > 0 \end{aligned} \tag{11}$$

- Consider this regression relationship within the Probit framework:
 - If $y_i^* \leq 0$ the individual does not consume the good ($z_i = 0$), but if $y_i^* > 0$ the individual consumes the good ($z_i = 1$).
 - We would estimate the this Probit as

$$\ln(L(\beta)) = \sum_{i=1}^N [z_i \ln(\Phi(x_i\beta)) + (1 - z_i) \ln(1 - \Phi(x_i\beta))] \tag{12}$$

Censored Regression – Continued

- Next, consider replacing the Probit function for $z_i = 1$ in Equation 12 with a normal distribution function from a regression

$$\ln(L(\beta)) = \sum_{i=1}^N \left[z_i \frac{1}{2} \left[\ln(2\pi) + \ln(\sigma^2) + \frac{(y - x_i\beta)^2}{\sigma^2} \right] + (1 - z_i) \ln \left[1 - \Phi \left(\frac{x_i\beta}{\sigma} \right) \right] \right] \quad (13)$$

Readings – Next Time

- * Greene, W.H. 2012. *Econometric Analysis*. Seventh Edition (Chapter 19 [Section 19.1 - 19.3 pp. 872 – 898])