Econometrics 6027
Lecture 4
Limited Dependent Variable Models

If the data beyond a threshold population value is being recoded to the threshold, or is just plain missing, we can use limited dependent variable models to estimate without bias.

In truncated data, the data beyond the limit is simply not observed. The relevant data is not available to the researcher. In censored data, the dependent variable is recoded to the value of the limit. The data gathering is defective.

For example, insurance data may be left truncated because people don't make claims for amounts which are below the cost of the deductible. In a censored sample, for example, individuals whose incomes fall short of some poverty line are assigned incomes equal to that poverty line. Truncated models address truncated data, and Tobit, Selectivity, and Double Hurdle models are used with censored data.

See Greene, Chapter 21 and Maddala, Chapter 6 for more analysis.

1 Truncated Samples

Consider:

- an observed dependent variable y
- a latent variable y^*
- a set of explanatory variables x
- lower threshold l
- upper threshold h

1.1 Observability Criteria:

Truncation from below:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > l, \\ \text{not observed} & \text{otherwise.} \end{cases}$$

Truncation from above:

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* < h, \\ \text{not observed} & \text{otherwise.} \end{cases}$$

Truncation from between:

$$y_i = \begin{cases} y_i^* & \text{if } l < y_i^* < h, \\ \text{not observed} & \text{otherwise.} \end{cases}$$

1.2 Theoretical considerations

With truncated samples, we don't observe data beyond a limit. If we just use OLS with this truncated data, our estimators will be biased, as we now show.

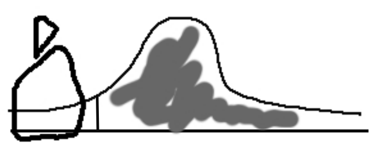
Consider the latent relationship:

$$y^* = x'\beta + u$$

Let $E(u|x) = 0$. Suppose that the observed dependent variable y is truncated from below at zero, and the researcher wants to estimate

$$y_i = x_i'\beta + u_i$$

If the researcher estimates this using the truncated data, the coefficients $\hat{\beta}$ will be biased. Visually,



The threshold to the right of D is $l - X_i\beta$. If we only use the shaded-in data, our estimates will ignore the D-part of the distribution. To unbias the distribution, we need to add back part D of the distribution.

This can also be seen using the conditional expectation underlying the regression:

$$\begin{aligned} E(y^*|x, y^* > 0) &= x'\beta + E(u|x, y^* > 0) \\ &= x'\beta + E(u|u > l - x'\beta) \\ &= x'\beta + \alpha f(l - x'\beta) \end{aligned}$$

Given that this is the conditional expectation, the correct model for estimation should have been:

$$y_i = x_i'\beta + \alpha f(l - x_i'\beta) + u_i$$

If we just use $y_i = x_i'\beta + u_i$, we're omitting the $\alpha f(l - x_i'\beta)$ and our estimates of $\hat{\beta}$ will be biased. Instead, we should use this correct model, including the $\alpha f(l - x_i'\beta)$ term. However, we can't estimate this using OLS because $f(x_i'\beta)$ is most likely non-linear on β.

Instead, we can use a maximum likelihood estimator. We need a density (PDF) that integrates to one across the support of y. To get the PDF to integrate to one we will scale the PDF by the probability that an observation falls in the observed region.

In terms of the graph above, we will weigh each observation we have (all of which are in the shaded zone) by the probability that an observation will be in the shaded zone instead of in "D". We do this by dividing the PDF by the CDF of the observed part of the distribution:

$$\begin{aligned} f(u|u > l) &= \frac{f(u)}{Pr(u > l)} \\ &= \frac{f(u)}{1 - Pr(u \leq l)} \text{ (by symmetry)} \\ &= \frac{f(u)}{1 - F(l)} \end{aligned}$$

Does this density integrate to 1, as any well-behaved probability should? Yes, it does.

$$\begin{aligned} \int_l^\infty f(u|u > l)\mathrm{d}u &= \int_l^\infty \frac{f(u)}{(1 - F(l))}\mathrm{d}u \\ &= \frac{1}{1 - F(l)} \int_l^\infty f(u)\mathrm{d}u \\ &= \frac{F(\infty) - F(l)}{1 - F(l)} \\ &= \frac{1 - F(l)}{1 - F(l)} = 1 \end{aligned}$$

1.3 Estimation of Truncated Model

Assume normality: $u \sim N(0, \sigma^2)$. We have:

$$f(u|u > l) = \frac{f(u)}{Pr(u > l)}$$

Where

$$\begin{aligned} Pr(u > l) &= Pr(\frac{u}{\sigma} > \frac{l}{\sigma}) \\ &= 1 - \Phi(\frac{l}{\sigma}) \\ &= \Phi(\frac{-l}{\sigma}) \end{aligned}$$

and where

$$\begin{aligned} f(u) &= \frac{1}{\sqrt{2\pi\sigma^2}} exp(\frac{-u^2}{2\sigma^2}) \\ &= \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} exp(\frac{-1}{2}(\frac{u}{\sigma})^2) \\ &= \frac{1}{\sigma}\phi(\frac{u}{\sigma}) \end{aligned}$$

So,

$$\begin{aligned} f(u|u>l) &= \frac{f(u)}{Pr(u>l)} \\ &= \frac{\frac{1}{\sigma}\phi(\frac{u}{\sigma})}{\Phi(\frac{-l}{\sigma})} \end{aligned}$$

The individual likelihood is:

$$\begin{aligned} L_i &= \frac{\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma})}{\Phi(\frac{x_i'\beta - l}{\sigma})} = \frac{\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma})}{1 - \Phi(\frac{l - x_i'\beta}{\sigma})} \text{ if } \sigma \neq 1, l \neq 0 \\ &= \frac{\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma})}{\Phi(\frac{x_i'\beta}{\sigma})} \text{ if } \sigma \neq 1, l = 0 \\ &= \frac{\phi(y_i - x_i'\beta)}{1 - \Phi(-x_i'\beta)} = \frac{\phi(y_i - x_i'\beta)}{\Phi(x_i'\beta)} \text{ if } \sigma = 1, l = 0 \end{aligned}$$

The total likelihood is:

$$\begin{aligned} L^T &= \prod L_i(\beta, \sigma) \\ \ln L^T &= \sum_{i=1}^{N} \ln L_i \end{aligned}$$

To summarize:

- we cannot use OLS
- so we use another method: Maximum Likelihood
- we have to use conditional density to avoid bias: we start with the individuals' conditional density
- to apply maximum likelihood, we need an assumption on the error: we assume normality
- we transform individual maximum likelihood of individual to the normal using the definition of the normal PDF

- we standardize by dividing by the standard deviation if $\sigma \neq 1$
- thereby obtaining the individual conditional normal likelihood
- we assume individual likelihoods are iid and that the product of them is thus the total likelihood.

1.4 Marginal Effects

Recall that the latent relationship is:

$$y^* = x'\beta + u$$

We want to find out the marginal effects from x_i:

$$\text{ME} = \frac{\partial E(y_i|y_i^* > 0)}{\partial x_i}$$

We start with the conditional expectation:

$$E(y_i|y_i^* > l) = x_i'\beta + E(u_i|u_i > l - x_i'\beta)$$

What is the last term, this conditional expectation of u?

$$\begin{aligned}
E(u_i|u_i > l - x_i'\beta) &= \int_{l-x_i'\beta}^{\infty} \frac{uf(u)}{(1 - F(l))}\mathrm{d}u \\
&= \frac{1}{1 - F(l - x_i'\beta)} \int_{l-x_i'\beta}^{\infty} uf(u)\mathrm{d}u \\
\text{Assuming normality,} & \\
&= \frac{1}{1 - \Phi(l - x_i'\beta)} \int_{l-x_i'\beta}^{\infty} u\phi(u)\mathrm{d}u
\end{aligned}$$

But we know that $\int_{l-x_i'\beta}^{\infty} u\phi(u)\mathrm{d}u = -\phi(u)$ since:

$$\begin{aligned}
\phi(u) &= \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}} \\
\frac{\partial\phi(u)}{\partial u} &= \frac{1}{\sqrt{2\pi}} e^{\frac{-u^2}{2}}(-u) \\
\frac{\partial\phi(u)}{\partial u} &= \phi(u)(-u) \\
-\frac{\partial\phi(u)}{\partial u} &= u\phi(u)
\end{aligned}$$

Thus

$$\int_{l-x_i'\beta}^{\infty} u\phi(u)\mathrm{d}u = \int_{l-x_i'\beta}^{\infty} -\frac{\partial\phi(u)}{\partial u}\mathrm{d}u = -\phi(u)|_{l-x_i'\beta}^{\infty}$$

Returning to the conditional expectation of u, we had:

$$\begin{aligned}
E(u_i|u_i > l - x_i'\beta) &= \frac{1}{1-\Phi(l-x_i'\beta)}\int_{l-x_i'\beta}^{\infty} u\phi(u)\mathrm{d}u \\
&= \frac{-\phi(u)\big|_{l-x_i'\beta}^{\infty}}{1-\Phi(l-x_i'\beta)} \\
&= \frac{-\phi(\infty)+\phi(l-x_i'\beta)}{1-\Phi(l-x_i'\beta)} \\
&= \frac{\phi(l-x_i'\beta)}{1-\Phi(l-x_i'\beta)} \\
&= \frac{\phi(-x_i'\beta)}{1-\Phi(-x_i'\beta)}(\text{if } l=0) \\
&= \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)} \text{ by symmetry of the distribution.}
\end{aligned}$$

This is the regressor that I was missing. Let's call it $\lambda = \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)}$. The conditional expectation of the y_i truncated at zero is then:

$$\begin{aligned}
[E(y_i|y_i^* > 0)] &= x_i'\beta + E(u_i|u_i > l - x_i'\beta) \\
&= x_i'\beta + \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)} \\
\text{If non-standard } &= x_i'\beta + \frac{1}{\sigma}\frac{\phi(\frac{x_i'\beta}{\sigma})}{\Phi(\frac{x_i'\beta}{\sigma})} \\
\text{if } l \neq 0 \qquad &= x_i'\beta + \frac{\phi(l)}{1-\Phi(l)}
\end{aligned}$$

Recall that marginal effects are:

$$\begin{aligned}
\text{ME} &= \frac{\partial E(y_i|y_i^* > 0)}{\partial x_i} \\
&= \frac{\partial(x_i'\beta + \lambda)}{\partial x_i} = \frac{\partial(x_i'\beta + \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)})}{\partial x_i} \\
&= \beta + \frac{\beta\frac{\partial\phi(x_i'\beta)}{\partial x_i}\Phi(x_i\beta) - \phi^2(x_i'\beta)\beta}{(\Phi(x_i\beta))^2} \\
&= \beta + \frac{\beta\phi'(x_i'\beta)\Phi(x_i'\beta) - \beta\phi(x'\beta)\Phi'(x\beta)}{\Phi^2(X'\beta)} \text{ since } \phi(x_i'\beta) = \Phi'(x_i'\beta).
\end{aligned}$$

Note that this is highly nonlinear in β. This is not asymptotic: if you estimate presuming normality, you will get β plus something. Exam questions could replace $l = 0$ with $l \neq 0$, make $\sigma \neq 1$, or use upper, lower, or both-ways truncation.

2 Censored Samples

Consider:

- an observed dependent variable y
- a latent variable y^*
- a set of explanatory variables x
- lower threshold l
- upper threshold h

2.1 Observability Criteria:

Censored from below:

$$y_i = \begin{cases} l & \text{if } y_i^* \leq l \\ y_i^* & \text{otherwise.} \end{cases}$$

Censored from above:

$$y_i = \begin{cases} h & \text{if } y_i^* \geq h \\ y_i^* & \text{otherwise.} \end{cases}$$

Censored from between:

$$y_i = \begin{cases} l & \text{if } y_i^* \leq l \\ y_i^* & \text{if } l < y_i^* < h, \\ h & \text{if } y_i^* \geq h. \end{cases}$$

A sample is censored if data are re-coded for a subset of the population. Unlike with truncated data, we DO observe the data beyond the threshold. The problem is rather that this data has been recoded to the value of the threshold.

Tobit, selectivity, and double-hurdle models are used to estimate censored data.

2.2 The Tobit Model

This model is attributable to Tobin, 1958. We have two types of observations: the uncensored, and the censored. The individual likelihood is the product of the probability of observing censored data and the probability of observing uncensored data:

$$L_i = \prod_{censored} P_c^{I=censored} \prod_{uncensored} P_{uc}^{I=uncensored}$$

Let's assume that our data is censored from below at $l = 0$. Then the probability of observing censored data is:

$$\begin{aligned} Pr(y_i = 0|x_i) &= Pr(y_i^* \leq 0|x_i) \\ &= Pr(x_i'\beta + u \leq 0|x_i) \\ &= Pr(u \leq -x_i'\beta|x_i) \\ &= 1 - F(x_i'\beta) \\ \text{Assuming standard normal,} \\ &= 1 - \Phi(x_i'\beta) \end{aligned}$$

The probability of observing an uncensored observation is:

$$\begin{aligned} Pr(y_i > 0|x_i) = f(y_i^* > 0) &= f(y_i - x_i\beta) \\ \text{Assuming normality,} \\ &= (\frac{1}{\sigma})\phi(\frac{y_i - x_i\beta}{\sigma}) \text{ if } u_i \sim N(0,\sigma) \\ &= \phi(y_i - x_i\beta) \text{ if } \sigma = 1 \end{aligned}$$

2.2.1 Estimation of Tobit Model

$$\begin{aligned} L_i &= P_{censored} * P_{uncensored} \\ L^T &= \prod_{y_i=0}(1 - \Phi(x'\beta))^{y_i=0} * \prod_{y_i\neq 0}\phi(y_i - x_i'\beta)^{y_i\neq 0} \\ \hat{\beta} &= argmax \ln L^T \end{aligned}$$

2.2.2 Marginal effects in the Tobit Model

The expected value of the observed dependent variable y_i censored from below at zero is

$$\begin{aligned} E(y_i|x_i) &= \Pr(y_i = 0|x_i)E(y_i|x_i, y_i < 0) + \Pr(y_i = y_i^*|x_i)E(y_i|x_i, y_i > 0) \\ &= (1 - \Phi(x_i'\beta)) * (0) + (\Phi(x_i'\beta))(x_i'\beta + \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)}) \end{aligned}$$

Where we computed that $E(y_i|x_i, y_i > 0) = (x_i'\beta + \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)})$for the truncated model. Then,

$$\begin{aligned} &= \Phi(x_i'\beta)\big(x_i'\beta + \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)}\big) \\ \text{ME} = \frac{\partial E(y_i|x_i)}{\partial x} &= \beta\Phi(x_i'\beta) + x_i'\beta^2\phi(x_i'\beta) + \beta\phi''(x_i'\beta) \\ &= \beta(\Phi(x_i'\beta) + x_i'\beta\phi(x_i'\beta) - \phi(x_i'\beta)) \\ \text{given that the derivative of } \phi(x_i'\beta) &= -\beta\phi(x_i'\beta) \end{aligned}$$

Analyzing the Tobit

- we could have endogenous thresholds: x such that l is $f(x)$ and x is independent of u.
- the Tobit Model treats the censored data as a corner solution, and gives it a value of zero. This will tend to weaken the coefficients, more so when a higher proportion of observations are censored.
- it's better to have one parameter/equation for censored data, and a different one for uncensored data, as with the selectivity model. The first one is a "participation" decision, the second one is for the dependent variable.

2.3 Selectivity Model

Consider:

$$y^* = x_i\beta + u_i$$
$$I_i^* = z_i\delta + v_i$$

In the selectivity model, the data is once again censored, but a latent indicator function, I_i^*, determines whether data is censored or not. This indicator function has its own explanatory variables, which help explain the value of I, that is, whether the data is censored or not. So we no longer treat the censored data as a corner solution, equalling zero; instead, we have a separate equation for the censored data.

For example, with left censoring, if $I_i^* > 0$, then we will observe $y = y^*$; if we observe $I_i^* \leq 0$ then y equals zero or whatever the censoring point is. The errors of the latent dependent variable equation and the latent indicator equation can be correlated:

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} \sim BVN \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}$$
$$cov(u_i v_i) = \rho = \frac{\sigma_{uv}}{\sigma_u \sigma_v}$$

Observability criteria:

Censored from below:

$$y_i = \begin{cases} y_i^* & \text{if } I_i^* > 0 \\ 0 & \text{otherwise. } (I_i^* \leq 0) \end{cases}$$

Or, put differently, $y_i = y_i^* * 1(I_i^* > 0)$

We have two types of observations: the uncensored, and the censored. The individual likelihood is the product of the probability of observing censored data and the probability of observing uncensored data. Assuming censoring from below at $l = 0$:

$$\begin{aligned} L_i &= Pr_{censored} * Pr_{uncensored} \\ &= Pr(I_i \leq 0|z_i)^{I(y_i=0)} * f(y_i|x_i, I_i > 0)^{I(y_i>0)} \end{aligned}$$

where the probability of the censored is:

$$\begin{aligned} Pr(y_i = 0|z_i) &= Pr(I_i^* \leq 0|z_i) \\ &= Pr(v_i \leq -z_i'\delta|z_i) \\ &= 1 - \Phi(z_i'\delta) \end{aligned}$$

and the conditional probability of the uncensored is:

$$f(y_i|x_i, I_i > 0) = f(y_i) * Pr(I_i^* > 0|y_i)$$

Where

$$f(y_i) = \frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma})$$

and where

$$\begin{aligned} Pr(I^* > 0|y_i) &= Pr(I_i^* > 0) = Pr(v_i > -z_i\delta) \\ &= \Phi\big(\frac{z_i'\delta + \rho(y_i - x_i'\beta)}{\sqrt{n - \rho^2}}\big) \\ \text{if } \sigma_{uv} = 0, & \\ &= \Phi(z_i'\delta) \end{aligned}$$

Thus,

$$f(y_i|x_i, I_i > 0) = \frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma}) * \Phi\big(\frac{z_i'\delta + \rho(y_i - x_i'\beta)}{\sqrt{n - \rho^2}}\big)$$

2.3.1 Estimation of Selectivity Model

The individual likelihood is thus:

$$\begin{aligned} L_i &= Pr_{censored} * Pr_{uncensored} \\ &= Pr(I_i \leq 0|z_i)^{I(y_i=0)} * f(y_i|x_i, I_i > 0)^{I(y_i>0)} \\ &= [1 - \Phi(z_i'\delta)]^{I(y_i=0)} * \left[\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma}) * \Phi\big(\frac{z_i'\delta + \rho(y_i - x_i'\beta)}{\sqrt{n - \rho^2}}\big)\right]^{I(y_i>0)} \end{aligned}$$

and the total likelihood is:

$$L_T = \prod_{y=0}[1 - \Phi(z_i'\delta)] * \prod_{y\neq 0}\left[\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma}) * \Phi\big(\frac{z_i'\delta + \rho(y_i - x_i'\beta)}{\sqrt{n - \rho^2}}\big)\right]$$

However, this is difficult to estimate by maximum likelihood methods. $\ln L^T$ would have to estimate δ and ρ jointly because of the correlation between u and v. To ease estimation of the model Heckman (1979) proposed a two-step technique that gives consistent but not efficient estimates.

Heckman Two-step Estimation of Selectivity

1^{st} stage: Estimate δ using a probit model for $I(\delta)$ on the entire sample

$$\hat{\delta}_{MLP} = \arg\max \ln \left(\prod_{y_i=0} (1 - \Phi(Z_i'\delta)) \prod_{y_i>0} \Phi(Z_i\delta) \right)$$

2^{nd} stage:
a. Using only the non-censored data (ie. as if it were a truncated subsample), construct:

$$\lambda(z_i\hat{\delta}) = \frac{\phi(z_i\hat{\delta})}{\Phi(z_i'\delta)}$$

recalling that

$$\begin{aligned} E(y_i|y_i > 0) &= x_i'\beta + E(u_i|I_i > 0) \\ &= x_i'\beta + \frac{\sigma_{uv}}{\sigma_u\sigma_v}\frac{\phi(z_i'\delta)}{\Phi(z_i'\delta)}, \text{ where the last term is the correction, } \lambda(z_i\hat{\delta}) \text{ in the notes} \end{aligned}$$

b. regress y_i on x_i and $\lambda(z_i\hat{\delta})$ by OLS on the truncated subsample:

$$y_i = x_i\beta + \alpha\lambda(z_i\hat{\delta}) + u_i$$

where $\lambda(z_i\hat{\delta})$ is a generated regressor, also known as an inverse mills ratio. This generates an unbiased estimate for $\hat{\beta}$.

Analyzing Selectivity

- we do this to avoid treating the zero as a corner solution
- the λ is a generated regressor which will affect the variance of the βs estimated in the last stage with this regressor. We can try to correct the standard errors, but it will still be inefficient.
- we are working with truncated data in the second step, not joint data, which affects standards
- We need to worry about the identification of the λ, also know as the hazard rate.

3 The Double Hurdle Model

This model was proposed by Cragg (1971) as an alternative to the selectivity model. Its name comes from the fact that there are two hurdles to be overcome before observing a non-censored observation.

Practical examples:
(1) Labour Supply:
  (a) do you want to work?
  (b) given that you choose to look for work, can you find a job?
(2) Credit constraints:
  (a) do you want to buy the good?
  (b) given that you want to buy the good, can you obtain a credit?

The Double Hurdle approach to the modelling of labour supply recognizes that there may be more than one reason why we observe a non-working individual in a sample, therefore a double hurdle model gives a richer characterization of a model of labour supply.

Consider:

$$\begin{aligned} y_i^* &= x_i'\beta + u_i \\ I_i^* &= z_i'\gamma + v_i \end{aligned}$$

Observability criteria:
Censored from below:

$$y_i = \begin{cases} y_i^* & \text{if } I_i^* > 0 \text{ and } y_i^* > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Put differently, $y_i = y_i^* * 1(I_i^* > 0 \text{ and } y_i^* > 0)$

The probability of observing a censored observation is:

$$\Pr(y_i = 0|x_i, z_i) = 1 - \Phi_2(z_i'\gamma, \frac{x_i'\beta}{\sigma_u}; \rho)$$

For non-censored observations,

$$\begin{aligned} \Pr(y_i > 0|x_i, z_i) &= \Pr(y_i^* > 0 \text{ and } I_i^* > 0) \\ &= \Pr(v_i > -z_i'\gamma \text{ and } u_i > -x_i'\beta) \\ &= \Phi_2(z_i'\gamma, \frac{x_i'\beta}{\sigma_u}; \rho) \end{aligned}$$

For non-censored observations, the likelihood contribution is identical to the Selectivity model.

3.0.2 Estimation of the Double Hurdle

Assuming no correlation ($\rho = 0$),

$$L(B, \gamma, \sigma_u) = \prod_{y_i=0} [1 - \Phi(z_i'\gamma)\Phi(\frac{x_i'\beta}{\sigma_u})] \prod_{y_i>0} \Phi(z_i'\gamma)\frac{1}{\sigma_u}\phi\left(\frac{y_i - x_i'\beta}{\sigma_u}\right)$$

4 Diagnostic and Specification Tests in QV and LDV Models

- a statistical framework
- Lagrange Multiplier (score) tests
- testing for normality in ML models
- testing for heteroskedasticity
- tests for omitted variables and regressor endogeneity
- an alternative LR testing framework

4.1 Introduction

Distributional assumptions underpin the specification of the majority of QV and LDV models. Indeed, most ML estimation is pre-conditioned on the assumption of normality. Naturally, the assumption ought to be tested. Similarly, we would like the ability to test for heteroskedasticity, omitted variables bias, and for possible endogeneity.

Some Background

Let

$$y_i = x_i'\beta + u_i$$

- Suppose we'd like to test the assumption that $u_i \sim N(0, \sigma^2)$.
- Take a sample of data $\{y_i, x_i\}$ for $i = 1, ..., n$,
- Likelihood function given k parameters $\theta = (\beta, \sigma)$ may be written

$$\begin{aligned} L(\theta) &= \sum_{i=1}^{n} \ln[\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma})] = \sum_{i=1}^{n} l_i \\ l_i &= \ln[\frac{1}{\sigma}\phi(\frac{y_i - x_i'\beta}{\sigma})] \end{aligned}$$

- Solve for ML estimates $\widetilde{\theta} = (\widetilde{\beta}, \widetilde{\sigma})$ using the following FOC;

$$
\begin{aligned}
\frac{\partial L(\widetilde{\theta})}{\partial \theta} &= \sum_{i=1}^{n} \frac{\partial l_i(\widetilde{\theta})}{\partial \theta} = \sum_{i=1}^{n} s_i(\widetilde{\theta}) \\
&= \sum_{i=1}^{n} \{ \frac{\partial l_i(\widetilde{\theta})}{\partial \theta_1}, \frac{\partial l_i(\widetilde{\theta})}{\partial \theta_2}, ..., \frac{\partial l_i(\widetilde{\theta})}{\partial \theta_k} \} = 0
\end{aligned}
$$

Example: a linear regression function

- Suppose that

$$y_i = x_i'\beta + u_i$$

and let $\varepsilon_i = \frac{u_i}{\sigma} = \frac{(y_i - x_i'\beta)}{\sigma}$

- Then,

$$l_i = \ln \left[\frac{1}{\sigma} \phi(\frac{y_i - x_i'\beta}{\sigma}) \right]$$

and

$$
\begin{aligned}
\frac{\partial l_i(\theta)}{\partial \beta} &= \sigma \phi(\cdot)^{-1} \frac{1}{\sigma} \frac{\partial [\phi(\cdot)]}{\partial \beta_1} \\
&= -\phi(\cdot)^{-1} \phi(\cdot) \frac{y_i - x_i'\beta}{\sigma} \frac{x_i}{\sigma}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial l_i(\theta)}{\partial \sigma} &= \sigma \phi(\cdot)^{-1} \left[\frac{1}{\sigma} \frac{\partial [\phi(\cdot)]}{\partial \sigma} - \phi(\cdot) \frac{1}{\sigma^2} \right] \\
&= \frac{1}{\sigma} [\varepsilon_i^2 - 1]
\end{aligned}
$$

- So, at the Maximum Likelihood estimates $\widetilde{\theta} = (\widetilde{\beta}, \widetilde{\sigma})$, the First Order conditions require that

$$
\begin{aligned}
N^{-1} \sum_{i=1}^{n} \widetilde{\varepsilon}_i &= 0 \\
N^{-1} \sum_{i=1}^{n} x_i \widetilde{\varepsilon}_i &= 0 \\
N^{-1} \sum_{i=1}^{n} \widetilde{\varepsilon}_i^2 &= 1
\end{aligned}
$$

where $\widetilde{\varepsilon}_i = (y_i - x_i'\widetilde{\beta})\widetilde{\sigma}^{-1}$.

4.2 A Lagrange Multiplier (LM) Testing Framework

- Under the null,

$$y_i = x_i'\beta + u_i$$

where $u_i \sim N(0, \sigma^2)$,

- Under an alternative,

$$y_i = x_i'\beta + z_i'\gamma + u_i$$

- Here, z_i represents an m-vector of additional regressors, and γ denotes the corresponding vector of parameters.
- Likelihood under the null is $L_0(\theta)$
- Likelihood under the alternative is $L_1(\theta, \gamma)$ for parameters $\theta = (\beta, \sigma)$ and γ,
- Hence, $L_0(\theta) = L_1(\theta, 0)$.
- The score test approach involves:
  – calculating $\frac{\partial L_1(\theta,0)}{\partial \gamma} = \frac{\partial L_1(\theta,0)}{\partial \gamma|\gamma=0}$
  – maximizing $L_0(\theta)$ obtain $\widehat{\theta}$
  – testing the joint significance of all m elements in
- LM test:

$$LM = i'S_A^*(S_A^{*'}S_A^*)^{-1}S_A^{*'}i$$

should be distributed as a chi-squared with m degrees of freedom under the null.

4.2.1 Testing for Normality

Statistically, the characteristics of the normal distribution are well-known. In particular, if we let ε represent a standard normally distributed random variable, then we may list properties of the *moments* $E(\varepsilon^j)$ as follows:

. $E(\varepsilon^1) = 0$(mean)
 $E(\varepsilon^2) = 1$ (variance)
. $E(\varepsilon^3) = 0$(skewness)
. $E(\varepsilon^4) = 3$ (kurtosis)

4.2.2 A Score Test for Normality

$$S_i^* = \left[\frac{\partial l_i\left(\widetilde{\widehat{\beta}}, \widetilde{\widehat{\sigma}}\right)}{\partial \widetilde{\widehat{\beta}}'}, \frac{\partial l_i\left(\widetilde{\widehat{\beta}}, \widetilde{\sigma}\right)}{\partial \widetilde{\sigma}}, \widetilde{\varepsilon}_1^3, (\widetilde{\varepsilon}_1^4 - 3) \right]$$

4.2.3 A Score Test for Hetero.

$$S_i^* = \left[\frac{\partial l_i\left(\widetilde{\widehat{\beta}}, \widetilde{\widehat{\sigma}}\right)}{\partial \widetilde{\widehat{\beta}}'}, \frac{\partial l_i\left(\widetilde{\widehat{\beta}}, \widetilde{\sigma}\right)}{\partial \widetilde{\sigma}}, \left(\widetilde{\varepsilon}_1^2 - 1\right) z_i' \right]$$

4.2.4 A Score Test for Omitted Variables

$$S_i^* = \left[\frac{\partial l_i \left(\widetilde{\widetilde{\beta}}, \widetilde{\widetilde{\sigma}} \right)}{\partial \widetilde{\widetilde{\beta}}'}, \frac{\partial l_i \left(\widetilde{\widetilde{\beta}}, \widetilde{\sigma} \right)}{\partial \widetilde{\sigma}}, \widetilde{\varepsilon}_1 z_i' \right]$$

4.3 Score Tests for Qualitative Variable Models

- When confronted with QV or LDV models, score tests aren't so easy.

- Why? One is unable to generate continuous realizations of the underlying disturbances or

residuals.

- Consider

$$y_i^* = x_i'\beta + \varepsilon_i$$

where $\varepsilon_i \sim N(0,1)$, with an observability criterion of the form

$$y_i = 1(y_i^* > 0)$$

Given ML estimates $\widetilde{\beta}$ but given the limited observability of y_i we can't solve for the residuals $\widetilde{\varepsilon}_i$

- We CAN produce 'estimates' of the residuals $\widetilde{\varepsilon}_i$ through knowledge of the fact that, for

$$y_i = 0, \ \widetilde{\varepsilon}_i < -x_i'\beta$$

- So, the *expected value* of ε_i conditional on $y_i = 0$ is

$$\begin{aligned} E\left(\varepsilon_i | y_i = 0\right) &= E\left(\varepsilon_i | \varepsilon_i < -x_i'\beta\right) \\ &= \frac{-\phi\left(x_i'\beta\right)}{1 - \Phi\left(x_i'\beta\right)}. \end{aligned}$$

- Similarly,

$$\begin{aligned} E\left(\varepsilon_i | y_i = 0\right) &= E\left(\varepsilon_i | \varepsilon_i > -x_i'\beta\right) \\ &= \frac{\phi\left(x_i'\beta\right)}{\Phi\left(x_i'\beta\right)}. \end{aligned}$$

- The *generalized error* ε_i (see Gourieoux, Montfort, Renault and Trognon, 1987) may therefore be defined as $E\left(\varepsilon_i | y_i\right)$, so

$$\begin{aligned} \varepsilon_i^{(1)} &= E\left(\varepsilon_i | y_i\right) \\ &= (1 - y_i) E\left(\varepsilon_i | y_i = 0\right) + y_i E\left(\varepsilon_i | y_i = 1\right) \\ &= (1 - y_i) E\left(\varepsilon_i | y_i < -x_i'\beta\right) + y_i E\left(\varepsilon_i | y_i > -x_i'\beta\right) \\ &= -(1 - y_i) \frac{\phi\left(x_i'\beta\right)}{1 - \Phi\left(x_i'\beta\right)} + y_i \frac{\phi\left(x_i'\beta\right)}{\Phi\left(x_i'\beta\right)}. \end{aligned}$$

- We generate generalized residuals $\widetilde{\varepsilon}$ (see Gourieoux, Montfort, Renault and Trognon, 1987) simply by replacing β with a ML estimate $\widetilde{\beta}$

- Higher order terms may be calculated using recursions
- Consider a Probit model, with an observation for which $y_i = 1$
- It can be shown in general that for $\varepsilon \sim N(0,1)$,

$$\begin{aligned}
\widetilde{\varepsilon}_i^{(2)} &= 1 - (x_i'\beta)\frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)} \\
\widetilde{\varepsilon}_i^{(3)} &= \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)}\left[2 + (x_i'\beta)^2\right] \\
\widetilde{\varepsilon}_i^{(4)} &= 3 - \frac{\phi(x_i'\beta)}{\Phi(x_i'\beta)}\left[3(x_i'\beta) + (x_i'\beta)^3\right]
\end{aligned}$$

- For $y_i = 0$, we take advantage of similar recursions to give;

$$\begin{aligned}
\varepsilon_i^{(1)} &= \frac{-\phi(x_i'\beta)}{1 - \Phi(x_i'\beta)} \\
\widetilde{\varepsilon}_i^{(2)} &= 1 + (x_i'\beta)\frac{\phi(x_i'\beta)}{1 - \Phi(x_i'\beta)} \\
\widetilde{\varepsilon}_i^{(3)} &= \frac{-\phi(x_i'\beta)}{1 - \Phi(x_i'\beta)}\left[2 + (x_i'\beta)^2\right] \\
\widetilde{\varepsilon}_i^{(4)} &= 3 - \frac{\phi(x_i'\beta)}{1 - \Phi(x_i'\beta)}\left[3(x_i'\beta) + (x_i'\beta)^3\right].
\end{aligned}$$

Substituting these residuals into the various test scores described earlier will permit testing binary response models.

4.4 An alternative Likelihood Ratio (LR) testing framework

These test are based on an auxiliary regression test and may be viewed as variable addition tests in which the significance of an extra set of regressors are tested by means of asymptotically likelihood ratio test statistics to resolve the hypothesis under investigation. They are easier to implement than the score tests.

4.4.1 LR test for non-normality in Probit models

Steps:

1. Estimate $y_i^* = x_i'\beta + \varepsilon_i$ to get ML estimates $\widetilde{\beta}$ and introduce them in $\log L_0$

2. Add test variables $\left(x_i\widetilde{\beta}\right)^2$ and $\left(x_i\widetilde{\beta}\right)^3$ to an auxiliary regression

$$y_i^* = x_i'\beta + \delta_1\left(x_i\widetilde{\beta}\right)^2 + \delta_2\left(x_i\widetilde{\beta}\right)^3 + \varepsilon_i$$

3. Obtain maximized log-likelihood $\log L_A$ from the auxiliary regression
4. The test statistic $2(\log L_A - \log L_0) \sim \chi^2_2$ under the null of normality.

4.4.2 LR test for heteroskedasticity in Probit models

Steps:

1. Estimate $y_i^* = x_i'\beta + \varepsilon_i$ to get ML estimates $\widetilde{\beta}$ and introduce them in $\log L_0$
2. Add test variables $\left(x_i\widetilde{\beta}\right) z_i$ to an auxiliary regression

$$y_i^* = x_i'\beta + \delta_1 \left(x_i\widetilde{\beta}\right) z_i + \varepsilon_i$$

   where z_i represent an $m-$vector of variables which may potentially cause the heteroskedasticity
3. Obtain maximized log-likelihood $\log L_A$ from the auxiliary regression
4. The test statistic $2(\log L_A - \log L_0) \sim \chi^2_m$ under the null of homoskedasticity.

4.4.3 LR test for heteroskedasticity in Probit models

Steps:

1. Estimate $y_i^* = x_i'\beta + \varepsilon_i$ to get ML estimates $\widetilde{\beta}$ and introduce them in $\log L_0$
2. Add test variables $\left(x_i\widetilde{\beta}\right) z_i$ to an auxiliary regression

$$y_i^* = x_i'\beta + \delta_1 z_i + \varepsilon_i$$

   where z_i represent an $m-$vector of variables which may potentially omitted
3. Obtain maximized log-likelihood $\log L_A$ from the auxiliary regression
4. The test statistic $2(\log L_A - \log L_0) \sim \chi^2_m$ under the null of no omission of variables.