Tobit/Selection

Nathaniel Beck

Department of Politics, NYU, New York, NY 10012, nathaniel.beck@nyu.edu

QII - Week 9

Intro - Censoring and Truncation

- Double Probit
  - Two dependent variables both observed
  - Example - State one arm/not, state two arm not
  - Simple with likelihood extension
  - Only matters if errors in two equations correlated
- Double probit with partial observability
  - Two probits, but only observe 1 if BOTH 1, otherwise 0
  - Takes two to tango
  - Example - IMF agreements
- Truncation
  - Only some observations get into the sample
  - Example - NJ Negative Income Tax Experiment - Only those with incomes under $10,000 in the experiment.
  - Only a problem if truncation is on the dependent variable

- Censoring (Selection)
  - All individuals (or a sample) are observed, but only some observations can be made on the dependent variables
  - Example - Wages of Woman (not PC! but won Heckman Nobel) - some don't work (for wages!) and so dependent variable is not observed
  - Spending on appliances - dependent variable not observed for those who don't buy (Tobit - but not why Tobin won Nobel)
  - Grades are not observed for those not admitted to college
  - Whether someone released on bail returns is not granted bail (Achen running example)
  - Bit more complicated but same idea - endogenous regimes
  - Example - countries choose an institution endogenously, then the institution affects outcomes

Simple double probit - two observed dvs

- Here we have two equations for two latent dependent variables
- BOTH observed only as probit (0 or 1)
- But related - one dv might be if state is failed, another if controlled by terrorists
- If two equations are independent, can estimate separately
- But may be related through error terms, a high draw on error for one may relate to high draw on second
- Discrete analog of SUR (seemingly unrelated regressions)

$$y_i^* = \mathbf{x}_i \beta_y + \epsilon_i^y \quad (1)$$
$$z_i^* = \mathbf{x}_i \beta_z + \epsilon_i^z \quad (2)$$
$$(3)$$

  where the ϵ's are BVN with zero mean, unit variance and arbitrary correlation
- No issue of simultaneity - x is exogenous
- Can have different iv's in each equation

- Each latent is converted to $0, 1$ in usual way, and all we observe is $0, 0$ or $0, 1$ or $1, 0$ or $1, 1$.
- Note that the likelihood of the four observed outcomes is just given by the integral of a BVN and so ml is easy
- If errors are highly correlated can make a difference
- Note that simple probit would be consistent, but is inefficient

Double probit with partial observability

- Model due to Poirrier, application in politics of Vreeland (and Prz.)
- Takes two to Tango, each actor makes a decision to enter an agreement (IMF)
- But only get an agreement if both say yes
- And only observe if agreement or not (have no idea who objected)
- Same model as above, but now only observe whether $1, 1$ or other
- Hence put in likelihood $P(1, 1)$ for $1, 1$ and one minus that for all others
- Advantage is different actors may have different IVs
- In theory, easier to think about what each actor wants, rather than just a probit on agreement or not

Truncation

- Suppose we observe data where

$$y_i = x_i'\beta + \epsilon_i, \epsilon_i \sim N(0, \sigma^2) \tag{4}$$

but that the sample is truncated (from below) at some value, a.

- a is exogenous
- For truncation, data simple does not get in the sample unless $y \geq a$.
- Thus, CONDITIONAL ON i BEING IN THE SAMPLE, ϵ is no longer normal but is distributed like a truncated normal. $y_i > a$ implies that $\epsilon_i > a - x_i'\beta$.
- Observations whose expected y is below a only get in the sample if they have a big error

- By the definition of conditional densities, a truncated density has

$$f(y|y > c) = \frac{f(y)}{1 - F(c)} \tag{5}$$

- Thus the density of observation i in our truncated sample is

$$\frac{\frac{1}{\sigma}\phi((y_i - x_i'\beta)/\sigma)}{1 - \Phi((a - x_i'\beta)/\sigma)} \tag{6}$$

- Assuming that a is a known constant we can thus do maximum likelihood fairly easily.

Censoring

- Censoring is more interesting
- Data is censored if all obs are in the data set, but only for some do we observe the dv we care about
- Suppose that, for Ms. i, that y is observed hours per week in the labor market. Let y^* represent the number of hours she desires to work. Then

$$y = 0 \text{ if } y^* \leq 0 \quad (7)$$
$$= y^* \text{ if } y^* > 0 \quad (8)$$

- Note that while Ms. i is observed to not be in the labor market, we do not know how many hours she desired to work
- Maybe her optimum is -20 (mine is!), but she has to solve a constrained optimization problem since one does not get to work less than zero hours

Tobit

- We observe how much people spend on an appliance
- y^* is desired spending
- But if $y^* < c_{min}$ where c_{min} is the minimum price of the appliance, you just observe they do not buy an appliance
- Observed y would then be zero, but that is not the desired y^*
- The tobit model (Tobin's probit, perhaps the most widely cited unpublished paper in the world, and not what Tobin won Nobel for!)

- $$y_i^* = x_i'\beta + \epsilon_i, \epsilon_i \sim N(0, \sigma^2) \tag{9}$$

$$y_i = 0 \text{ if } y^* \leq 0 \tag{10}$$
$$= y^* \text{ if } y^* > 0 \tag{11}$$

- How do we setup the likelihood?
- For the censored values, it is just P(being censored), that is,

$$P(y_i^* \leq 0) = P(x_i'\beta \leq -\epsilon_i) = 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right) \tag{12}$$

- while for the non-censored values it is just a normal density, $\phi(y_i; x_i'\beta, \sigma^2)$.

$$L = \prod_{y_i > 0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(y_i - x_i'\beta)^2}{2\sigma^2}} \prod_{y_i \leq 0} 1 - \Phi\left(\frac{x_i'\beta}{\sigma}\right) \tag{13}$$

  which can be maximized fairly easily

Issues

- If the censoring point is not 0 but must be estimated, use

$$\min(y_i | y_i > 0) \tag{14}$$

  This will exceed the true censoring point, but is better than using 0
- Can easily do Tobit where top values are censored - cannot spend more than some amount on an appliance, though probably not true today!
- Can censor at top and bottom
- All easy conceptually and in Stata
- In practice top censoring probably of less interest than bottom, but not always
- Why not use OLS on the uncensored obs?
- Consider the draws of ϵ for i's with low $x_i'\beta$. They must, on average, be positive.
- But for those with high $x_i'\beta$ can have any ϵ. The x and ϵ not independent, so OLS not good.

Interpretation

- How does Tobit arise?
- Constrained maximization
- When is zero really zero and not just that what you want to spend is less than you can spend?
- What if you just have a bunch of zero's in your data????

Selection bias

- Original econometrics due to Heckman
- Selection bias models have two equations. One is for whether an individual is selected or not ($s_i = 0$ or 1)
- Second is outcome equation (y_i, which is scored continuously, though it could easily be made into a quantal variable) which gives the outcome (wages, grades) for those who are selected. item Why not just estimate outcome model for those who are selected.
- Suppose the error terms in both equations are related (they share a common omitted variable).
- In terms of college admissions/outcomes this might be some measures of ability that are not in the analyst's record.
- Then you are more likely to get selected with a positive 'error' and since errors over the two equations are correlated the usual OLS assumptions about the error process for the outcome equation are incorrect.
- (OLS is fine if the error terms are independent.)

Model setup

- Selection equation

$$s_i^* = w_i'\gamma + \mu_i \quad (15)$$
$$s_i = 1 \text{ if } s_i^* > 0 \quad (16)$$
$$s_i = 0 \text{ if } s_i^* \leq 0 \quad (17)$$

- Outcome equation

$$y_i = x_i\beta + \epsilon_i, \text{ Observed if } s_i = 1 \quad (18)$$

- Stochastic Model

$$(s_i^*, y_i) \sim N(w_i'\gamma, x_i'\beta, 1, \sigma_\epsilon^2, \rho) \quad (19)$$
$$(\mu_i, \epsilon_i) \sim N(0, 0, 1, \sigma_\epsilon^2, \rho) \quad (20)$$

- (Why is the variance of $\mu = 1$ and the censoring point 0. For the same reason that we can't estimate cutoff and variance of a probit!)

Moments of BVN

- Let y and z be bivariate normal with correlation ρ.
-

$$E(y|z > a) = \mu_y + \rho\sigma_y\lambda(\alpha_z) \quad (21)$$
$$Var(y|z > a) = \sigma_y^2(1 - \rho^2\delta(\alpha_z)) \quad (22)$$

where

$$\alpha_z = \frac{a - \mu_z}{\sigma_z} \quad (23)$$
$$\lambda(\alpha_z) = \frac{\phi(\alpha_z)}{1 - \Phi(\alpha_z)} \text{ 'Inverse Mills Ratio'} \quad (24)$$
$$\delta(\alpha_z) = \lambda(\alpha_z)(\lambda(\alpha_z) - \alpha_z) \quad (25)$$

Heckman two step

- This gives an easy (non-ml) way to estimate these models, due to Heckman.
  - Do probit to estimate $\hat{\gamma}$.
  - Use this to calculate $\hat{\lambda}_i = \frac{\phi(w_i'\hat{\gamma}}{\Phi(w_i'\hat{\gamma}}$.
  - Estimate β in outcome equation by running a regression of y on x and $\hat{\lambda}$.
- Term added is called the IMR (inverse Mills ratio)
- Not fully efficient, but corrects mean, so consistent and unbiased
- Matters most if there is a good selection equation
- Note that if same iv's are in the selection and outcome equations, than it will be the case that the IMR is close to being colinear with the x's (it is just a non-linear fn of them)
- So get best results when there is some variable(s) in the selection equation not in the outcome equation
- But what?

Likelihoods

- Censored Observations
- For a censored individual all we know is that $s_i^* \leq 0$ so its contribution to the likelihood is

$$L_i^c = P(s_i^* \leq 0) \tag{26}$$
$$= P(w_i'\gamma + \mu_i \leq 0) \tag{27}$$
$$= P(\mu_i \leq -w_i'\gamma) \tag{28}$$
$$= \Phi(-w_i'\gamma) \tag{29}$$

-
$$L_i^o = \frac{1}{\sigma_\epsilon}\phi\left(\frac{y_i - x_i'\beta}{\sigma_\epsilon}\right) \quad (30)$$
$$\Phi\left(\frac{w_i'\gamma + \frac{\rho}{\sigma_\epsilon}(y_i - x_i'\beta)}{\sqrt{1-\frac{\rho}{\sigma_\epsilon}}}\right) \quad (31)$$

- Thus the log likelihood of the sample is

$$\ln L = \sum_{i=1}^{n} s_i \ln L_i^o + (1-s_i)\ln L_i^c \quad (32)$$

Double Probit with selection

- Suppose that the outcome equation is also a probit

$$y_i^* = x_i'\beta + \epsilon_i \quad (33)$$
$$y_i = 1 \text{ if } y_i^* > 0 \quad (34)$$

where we now (as in probit) set $\sigma_\epsilon^2 = 1$.
- The likelihood for a censored observation is as above (Equation 29).
- To simplify notation, let

$$\eta_i = x_i'\beta \quad (35)$$
$$\zeta_i = z_i'\gamma \quad (36)$$

- For uncensored observations the contribution to the likelihood is

$$L_i^o = \frac{1}{2\pi\sqrt{1-\rho^2}} \quad (37)$$
$$\int_{-\infty}^{\eta_i}\int_{-\infty}^{\zeta_i} e^{-\left[\frac{1}{2(1-\rho^2)}(\eta_i^2 - \rho\eta_i\zeta_i + \zeta_i^2)\right]} d\eta_i d\zeta_i \quad (38)$$

with the log likelihood being then formed as for the previous case

Endogenous regimes

- Consider a model which has a probit for whether units choose to be in regime 1 or regime 2 (could be IMF agreement or not, or in econ example, farming or fishing)
- Once in the chosen regime, there is an outcome.
- But note that people choose regimes that are good for them, so errors in selection equation are related to errors in outcome equation
- Looks just like a selection nodel, but contribution of each observation to the likelihood is just a single integral
- Moral - you can't just type selbias in Stata, you need to think about the model

Matching

- Heckman is selection on unobservables
- Problem it solves is that things we cannot observe (how much somoene expects to earn) conditions both whether they enter the labor force AND how much they make
- Matching tries to solve by matchig units on observables
- Regression as a form of matching
- Propensity scores