ECON 312

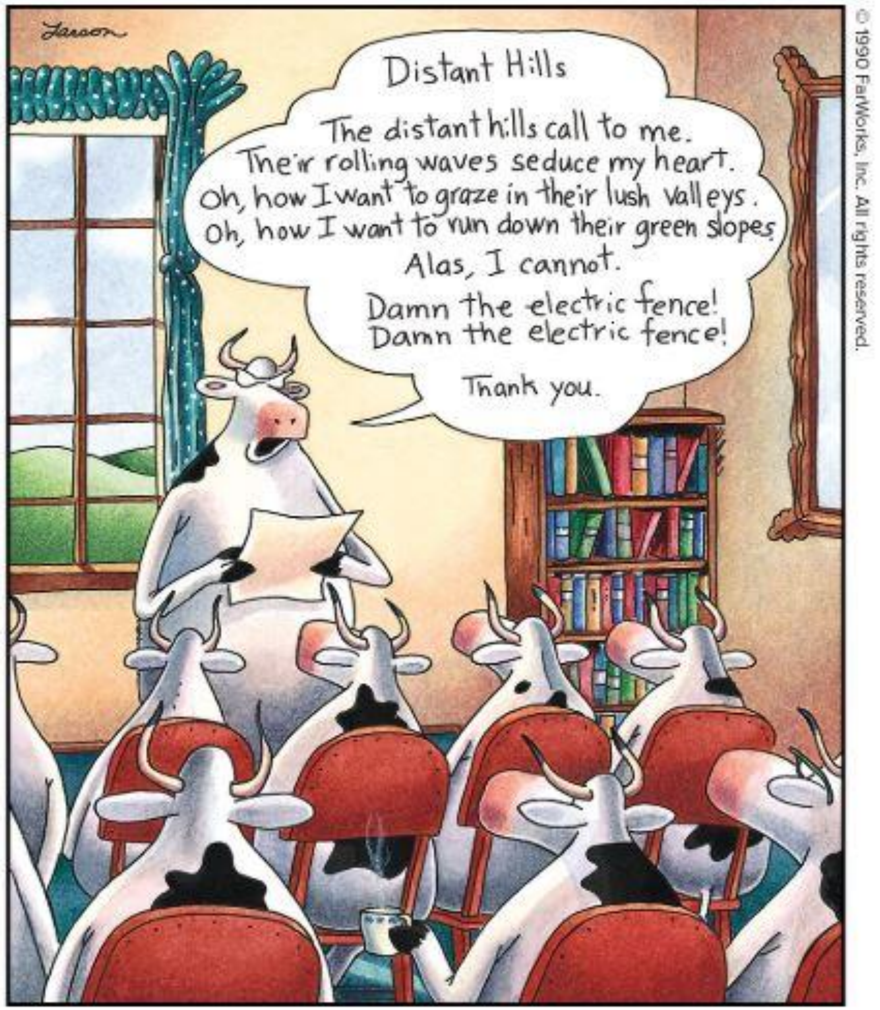
Monday, April 27

Models with Restricted Dependent Variables

Readings: Wooldridge, Section 17.2, 17.4, 17.5

Class notes: 164 - 171

Today’s Far Side offering



Cow poetry

A little poetry for your cultural enlightenment…

Context and overview

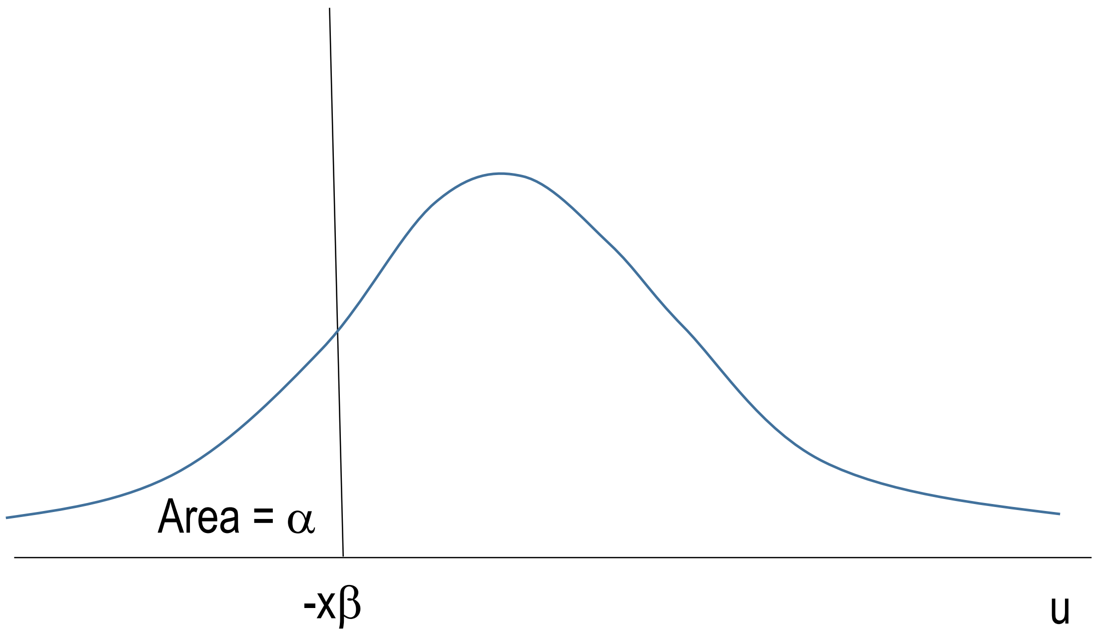
- The final set of limited-dependent-variable models we consider have a dependent variable that is continuous but has **restricted range**
  - Variables that cannot be negative are the most common application
  - There can also be upper bounds either on a variable's feasible values or on the range of observable outcomes
- We consider four sets of estimators:
  - **Tobit** estimators for corner solutions
  - **Censored regression:** we observe the limit value for extreme observations
  - **Truncated regression:** we observe neither x nor y for observations beyond the limit
  - **Incidental truncation** where the criterion for truncation depends on the error

Some examples

- Corner solution: many people choose to consume zero (tobit)
- Top-coded surveys: often a maximum feasible response though actual value is higher (censored regression)
- Sold-out events: observed quantity cannot be larger than capacity, though demand might be (censored regression)
- Our sample of y and x only contains people with limited y range, though others exist (truncated regression)
- Our ability to observe y and x is correlated with u, as with both wage and hours missing for non-workers (incidental truncation: Heckit)

Corner solutions: The tobit model

- $y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i$ cannot be negative, so u_i cannot be $< -\mathbf{x}_i\boldsymbol{\beta}$
- Depending on $\mathbf{x}_i$, some share α of probability falls in negative region, where person chooses 0
- Distribution of u_i is hybrid:
  - Continuous for values $> -\mathbf{x}_i\boldsymbol{\beta}$
  - Discrete: $\Pr[u = -\mathbf{x}_i\boldsymbol{\beta}] = \alpha$
- This is called a **censored distribution**



What's wrong with simple estimators?

- **OLS**?
  - Leads to predictions of negative y
  - Ignores diversity of $\mathbf{x}_i$ values leading to $y = 0$
- Just use **non-censored observations** with $y > 0$?
  - Violates random-sample assumption
  - Observations with negative u get left out more often
- Use **ln*y*** as dependent variable?
  - Not defined for $y_i = 0$
  - (Sometimes use $\ln(y_i + 1)$ as alternative)

Tobit model

- Latent variable $y_i^* = \mathbf{x_i\beta} + u_i$ with normal u_i and observable dependent variable
$$y_i = \begin{cases} y_i^*, \text{ if } y_i^* \geq 0, \\ 0, \text{ otherwise} \end{cases}$$
- Conditional density is censored distribution:
$$f\left(y_i \mid \mathbf{x_i}\right) = \frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x_i\beta}}{\sigma}\right) \text{ for observed } y_i > 0$$
$$\Pr\left[y_i = 0 \mid \mathbf{x_i}\right] = 1 - \Phi\left(\mathbf{x_i\beta}/\sigma\right)$$
- Tobit likelihood function (max over $\boldsymbol{\beta}$ and σ):
$$\ln L\left(\boldsymbol{\beta}, \sigma; y, x\right) = \sum_{i:y_i=0} \ln\left[1 - \Phi\left(\frac{\mathbf{x_i\beta}}{\sigma}\right)\right] + \sum_{i:y_i>0} \ln\left[\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x_i\beta}}{\sigma}\right)\right]$$

Specifying tobit estimator

- Can have left-censored, right-censored, or doubly-censored distributions
  - Example of last: share of pinot noir in individual's wine consumption is bounded by zero below and one above
- Censorship bounds must be specified in advance
  - Stata will by default choose sample maximum/minimum values
- **tobit *depvar indvars* , ll(0)** does tobit model with dependent variable *depvar* left-censored at zero
  - **ul( )** option can specify upper limit

Tobit expectations of *y*

- Conditional on $y > 0$ expectations

$$E\left[\left(y_i \mid \mathbf{x_i}\right) \mid y_i > 0\right] = E\left[y_i \mid y_i > 0, \mathbf{x_i}\right] = \mathbf{x_i}\boldsymbol{\beta} + \sigma \frac{\phi\left(\dfrac{\mathbf{x_i}\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\dfrac{\mathbf{x_i}\boldsymbol{\beta}}{\sigma}\right)}$$

$$= \mathbf{x_i}\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{x_i}\boldsymbol{\beta}}{\sigma}\right), \quad \text{where "inverse Mills ratio" is } \lambda(c) \equiv \frac{\phi(c)}{\Phi(c)}$$

- Unconditional expectation of *y*

$$E\left[y_i \mid \mathbf{x_i}\right] = \Phi\left(\frac{\mathbf{x_i}\boldsymbol{\beta}}{\sigma}\right)\mathbf{x_i}\boldsymbol{\beta} + \sigma\phi\left(\frac{\mathbf{x_i}\boldsymbol{\beta}}{\sigma}\right)$$

Partial effects in tobit

- Effect of x_j on **conditional expectation**:

$$\frac{\partial E[y_i \mid y_i > 0, \mathbf{x_i}]}{\partial x_j} = \beta_j \left\{ 1 - \lambda\left(\frac{\mathbf{x_i\beta}}{\sigma}\right)\left[\frac{\mathbf{x_i\beta}}{\sigma} + \lambda\left(\frac{\mathbf{x_i\beta}}{\sigma}\right)\right]\right\}$$

- Effect of x_j on **unconditional expectation**:

$$\frac{\partial E[y_i \mid \mathbf{x_i}]}{\partial x_j} = \beta_j \Phi\left(\frac{\mathbf{x_i\beta}}{\sigma}\right)$$

Post-estimation tobit Stata commands

- **Predicted values**:
  - predict , pr(0, .) gives probability that each observation is not censored
  - predict , e(0, .) gives predicted *y* value conditional on not censored
  - predict , ystar(0, .) gives unconditional predicted *y*
- **Marginal effects**:
  - margins dydx(*) , pr(0, .) gives effects of all variables on probability that observation is not censored
  - margins dydx(*) , e(0, .) gives effects of all variables on expected value conditional on it being not censored
  - margins dydx(*) , ystar(0, .) gives effects of all variables on unconditional expected value

Censored regression

- We know x for all observations, but observe $y = c$ for values past c
- Different from tobit because y can actually be larger/smaller than c, but we observe c if it is
- Censorship is not part of choice, just an error in our y observation
- Likelihood function is same as tobit; so is estimator
- Simpler interpretation: β_j is effect of x_j on $E(y \mid x)$, period
- Use Stata tobit command, but don't bother with margins afterward to compute effects

Truncated regression

- Like censored regression, except neither y nor x can be observed for extreme observations
- Incidental truncation: **Heckman's "heckit" estimator**
  - Consider labor-supply equation where y = hours worked and the key regressor of interest is the wage offer
  - Only observe wage offer for people who work
  - People with low u are less likely to work
  - Work status (variable determining truncation) is correlated with u
  - OLS is biased because of non-random sample selection over u
    - OLS would be fine if sample selection criterion were not correlated with u

Heckit regression

- Details in notes or Wooldridge 17-5b
- Two-step estimation procedure:
  1. Sample-selection equation: Probit model to determine inclusion in sample based on available variables *z* (which must include at least one variable not in the regression … like an instrument)
  2. Corrected regression adding inverse Mills ratio for each observation as an additional regressor

$$y_i = \mathbf{x_i}\boldsymbol{\beta} + \rho\hat{\lambda}_i + u_i \text{ with } \hat{\lambda}_i = \lambda\left(\mathbf{z_i}\hat{\boldsymbol{\gamma}}\right) = \frac{\phi\left(\mathbf{z_i}\hat{\boldsymbol{\gamma}}\right)}{\Phi\left(\mathbf{z_i}\hat{\boldsymbol{\gamma}}\right)}$$

- Stata command heckman

Review and summary

- We have examined three major models for dependent variables that have restricted range
- Tobit handles corner solutions, where individuals choose the limit value
    - Marginal effects are complex, with multiple cases of interest
- Censored regression is simpler because it involves only inability to observe extreme values, not inability to choose them
- Incidental truncation occurs when we can't observe either y or x for extreme observations and the criteria for truncation are related to the error term
    - Heckit two-step estimator is appropriate for this model

Another bad economist joke …

Three guys decide to play a round of golf: a priest, a psychologist, and an economist. They get behind a very slow twosome, who, despite having caddies, are taking all day to line up their shots and then four-putting every green. By the 8th hole, the three men are complaining loudly about the slow play ahead of them and swearing up a storm.

The priest says, "Holy Mary, I pray that they should take some lessons before they play again." The psychologist says, "I swear there are people who like to play golf slowly." The economist says, "I didn't expect to spend this much time playing a round of golf."

By the 9th hole, they have had it with slow play. The psychologist goes up to a caddie and demands that they be allowed to play through. The caddie says that would be fine, and explains that the two golfers are blind, and that both are retired firemen who lost their eyesight saving people in a fire. This explains their slow play, states the caddie. "Would you please not swear and complain so loudly?"

The priest is mortified, saying, "Here I am, a man of the cloth, and I've been swearing to the slow play of two blind men." The psychologist is also mortified, saying, "Here I am, a man trained to help others with their problems, and I've also been complaining about the slow play of two blind men."

The economist ponders the situation. He goes back to the caddies and asks, "Listen, the next time they play, could it be at night?"

--Taken from Jeff Thredgold, *On the One Hand: The Economist's Joke Book*

What's next?

- Good question!
- As of the preparation of this lecture, I have received no requests on the set of extended topics in the final section of the reading list
- Unless I hear strong preferences by the weekend, we will do a brief discussion of **multiple-imputation models for missing data** and the **quantile regression** model