Lecture 5

Limited Variables 4 of 5:
Tobit & Poisson Models

Michael Curran

Trinity College Dublin

JS Econometrics

Lecture 5 Outline

Tobit Model

Corner Solutions

Zero for nontrivial fraction of population but roughly continuously distributed over positive values. Example? Let y be continuous over strictly positive values but take on zero with positive probability. Because distribution of y piles up at zero, y can't be a conditionally Normal distribution. So all inference would only have asymptotic justification like LPM. Overcomes limitations of LPM. **Tobit model** is convenient and typically expresses the observable response y in terms of an underlying latent variable:

$$y^* = \beta_0 + \mathbf{x}\boldsymbol{\beta} + u, \quad u|\mathbf{x} \sim N(0, \sigma^2) \tag{1}$$

$$y = max(0, y^*) \tag{2}$$

Latent variable y^* satisfies CLRM assumptions: Normal, homoscedastic distribution with a linear conditional mean. (2) implies observable variable y is y^* when $y^* \geq 0$ but $y = 0$ when $y^* < 0$.

Tobit Model

Corner Solutions

Since y^* is Normal, y has continuous distribution over strictly positive values. In particular, density of y given $\mathbf{x}$ is same as density of y^* given $\mathbf{x}$ for positive values. Further:

$$P(y=0|\mathbf{x}) = P(y^* < 0|\mathbf{x}) = P(u < -\mathbf{x}\boldsymbol{\beta}|\mathbf{x})$$
$$= P(\frac{u}{\sigma} < -\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}|\mathbf{x}) = \Phi(-\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}) = 1 - \Phi(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma})$$

since $\frac{u}{\sigma}$ is N(0,1) and is independent of $\mathbf{x}$; intercept is in $\mathbf{x}$. If $(\mathbf{x_i}, y_i)$ is random draw from population, density of y_i given $\mathbf{x}_i$ is given by:

$$(2\pi\sigma^2)^{\frac{-1}{2}} \exp\left[\frac{-(y-\mathbf{x}_i\boldsymbol{\beta})^2}{(2\sigma)^2}\right] = \frac{1}{\sigma}\phi\left[\frac{(y-\mathbf{x}_i\boldsymbol{\beta})}{\sigma}\right] \quad y > 0 \quad (3)$$

$$P(y_i = 0|\mathbf{x}_i) = 1 - \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right) \quad (4)$$

Tobit Model

Corner Solutions

From (3) & (4) we can obtain the log-likelihood function for each observation i:

$$\ell_i(\boldsymbol{\beta}, \sigma) = 1(y_i = 0) \log \left[1 - \Phi\left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{\sigma}\right)\right] + 1(y_i > 0) \log \left\{\frac{1}{\sigma}\phi\left[\frac{(y_i - \mathbf{x}_i \boldsymbol{\beta})}{\sigma}\right]\right\} \tag{5}$$

The log-likelihood for a random sample of size n is obtained by summing (5) across all i.

MLE of $\boldsymbol{\beta}$ and σ require numerical methods (mostly done easily on packaged routine). The matrix formula for SE is complicated.

Testing multiple exclusion restrictions: Wald test (similar form to logit/probit case) / LR test (as in logit/probit but with Tobit log-likelihood for restricted and unrestricted models).

Question 17.3.

Tobit Model

Interpreting Estimates

Cannot interpret $\hat{\beta}_j$ from Tobit MLE the same as from linear model OLS.

From (1), β_j measure partial effects of x_j on $E(y^*|\mathbf{x})$. Want to explain y (observed outcome, e.g. hours worked). Can estimate $P(y=0|\mathbf{x})$ from (4) and this allows us est $P(y>0|\mathbf{x})$ but what if want to estimate expected value of y as function of $\mathbf{x}$?

With Tobit, 2 expectations are interesting: (i) $E(y|y>0,\mathbf{x})$ 'conditional expectation' and (ii) $E(y|\mathbf{x})$ unfortunately called 'unconditional expectation'. Given $E(y|y>0,\mathbf{x})$, can easily find $E(y|\mathbf{x})$:

$$E(y|\mathbf{x}) = P(y>0|\mathbf{x}) \cdot E(y|y>0,\mathbf{x}) = \Phi(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}) \cdot E(y|y>0,\mathbf{x}) \tag{6}$$

Tobit Model

Interpreting Estimates

To get $E(y|y>0,\mathbf{x})$ use result:

$z \sim N(0,1) \implies E(z|z>c) = \frac{\phi(c)}{[1-\Phi(c)]}$ for any constant c But:

$$\begin{aligned} E(y|y>0,\mathbf{x}) &= \mathbf{x}\boldsymbol{\beta} + E(u|u>-\mathbf{x}\boldsymbol{\beta}) \\ &= \mathbf{x}\boldsymbol{\beta} + \sigma E[\frac{u}{\sigma}|\frac{u}{\sigma} > -\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}] \\ &= \mathbf{x}\boldsymbol{\beta} + \frac{\sigma\phi(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma})}{\Phi(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma})} \end{aligned}$$

since $\phi(-c) = \phi(c)$, $1-\Phi(-c) = \Phi(c)$ and $\frac{u}{\sigma}$ has standard Normal distribution independent of $\mathbf{x}$. So

$$E(y|y>0,\mathbf{x}) = \mathbf{x}\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right) \tag{7}$$

where $\lambda(c) = \frac{\phi(c)}{\Phi(c)}$ is called the **inverse Mills ratio**.

Tobit Model

Interpreting Estimates

Since $\Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)=\phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)$, combining (6) & (7) gives:

$$E(y|\mathbf{x})=\Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\left[\mathbf{x}\boldsymbol{\beta}+\sigma\lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\right]=\Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\mathbf{x}\boldsymbol{\beta}+\sigma\phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right) \quad (8)$$

With $\boldsymbol{\beta}$ estimates we can be sure that predicted values for y (estimates of $E(y|\mathbf{x})$) are positive, but this comes at a cost: equation (8) is more complicated than a linear model. Partial effects of x_j on $E(y|y>0,\mathbf{x})$ and $E(y|\mathbf{x})$ have same sign as coefficient β_j but magnitude of effects depend on values of *all* explanatory variables and parameters. Since σ appears in (8), the partial effects depend on σ also.

Tobit Model

Interpreting Estimates

Partial effects: let x_j be continuous and assume x_j is not related to other regressors. Then:

$$\frac{\partial E(y|y>0,\mathbf{x})}{\partial x_j} = \beta_j + \beta_j \cdot \underline{d\lambda}\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)$$

Differentiate $\lambda(c) = \frac{\phi(c)}{\Phi(c)}$ and use $\frac{d\Phi}{dc} = \phi(c)$ and $\frac{d\phi}{dc} = -c\phi(c)$.

ICBST: $\frac{d\lambda}{dc} = -\lambda(c)[c + \lambda(c)]$ So we get:

$$\frac{\partial E(y|y>0,\mathbf{x})}{\partial x_j} = \beta_j \left\{1 - \lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\left[\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma} + \lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\right]\right\} \tag{9}$$

Estimate (9) by plugging in MLEs of β_j and σ. Like logit/probit, plug in values for x_j like mean values or other interesting values.

Subtlety: σ appears in partial effects directly so crucial to estimate this for estimating partial effects; σ called 'ancillary' parameter (i.e. auxiliary or important – misleading terminology).

Usual economic quantities can be computed, e.g. elasticities.

Tobit Model

Interpreting Estimates

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \frac{\partial P(y>0|\mathbf{x})}{\partial x_j} \cdot E(y|y>0,\mathbf{x}) + P(y>0|\mathbf{x}) \cdot \frac{\partial E(y|y>0,\mathbf{x})}{\partial x_j} \tag{10}$$

$$\frac{\partial P(y>0|\mathbf{x})}{\partial x_j} = \frac{\beta_j}{\sigma}\phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right) \tag{11}$$

Thus can estimate each term in (10) once we plug in MLEs of β_j and σ and particular values of the x_j.

$$\frac{\partial E(y|\mathbf{x})}{\partial x_j} = \beta_j \Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right) \tag{12}$$

(12) allows us to roughly compare OLS and Tobit estimates. OLS slope coefficients (e.g. $\hat{\gamma}_j$ from regression of y_i on $x_{i1}, x_{i2}, \ldots, x_{ik}$ where $i = 1, \ldots, n$ i.e. using all of data) are direct estimates of $\partial E(y|\mathbf{x})/\partial x_j$. To make Tobit coefficients $\hat{\beta}_j$ comparable to $\hat{\gamma}_j$, we multiply $\hat{\beta}_j$ by an adjustment factor.

Tobit Model

Interpreting Estimates

2 approaches for computing an adjustment factor:

1. Evaluate $\Phi\left(\frac{\mathbf{x}\hat{\beta}}{\hat{\sigma}}\right)$ at sample averages to obtain $\Phi\left(\frac{\bar{\mathbf{x}}\hat{\beta}}{\hat{\sigma}}\right)$
2. Average the individual adjustment factors $n^{-1}\sum_{i=1}^{n}\Phi\left(\frac{\bar{\mathbf{x}}_i\hat{\beta}}{\hat{\sigma}}\right)$

For comparing scaled Tobit coefficients to OLS coefficients, second scale factor generally is more appropriate. Both scale factors will tend to be closer to one when there are relatively few observations with $y_i = 0$. In extreme case that all $y_i > 0$, Tobit and OLS estimates are identical. With discrete explanatory variables, comparing OLS and Tobit est is not so easy; however, scale factor for continuous explanatory variables is often a useful approximation.

Tobit Model

Specification Issues

- Tobit (especially (7) and (8)) crucially relies on Normality and homogeneity in underlying latent variable model, unlike linear model where Normality of y does not affect unbiasedness / consistency / large sample inference and heterogeneity does not affect unbiasedness / consistency of OLS and we can use robust SE for inference.
- Tobit: if any assumptions in (1) fail, it's hard to know what we're estimating but for moderate departures, Tobit model is likely to provide good estimates for partial effects on conditional means.
- Can allow more general assumptions in (1) but hard to estimate and interpret.
- Limitation: expected value conditional on $y > 0$ linked with probability $y > 0$.

Tobit Model

Specification Issues

How can we informally evaluate whether Tobit is appropriate? Estimate probit where binary outcome $w = 1$ if $y > 0$ and $w = 0$ if $y = 0$. Then from (4), w follows probit model where coefficient on x_j is $\gamma_j = \frac{\beta_j}{\sigma}$ so we can estimate ratio of β_j to σ by probit for each j. If Tobit model holds, probit estimate $\hat{\gamma}_j$ should be 'close' to $\hat{\beta}_j/\hat{\sigma}$ where these are Tobit estimates. Never identical (due to sampling error) but warning signs include different signs or same signs and much different magnitudes; sign changes or magnitude differences on explanatory variables that are insignificant in both models should not be a cause for concern. If Tobit is inappropriate, *hurdle* or *two-part* models can be used and have the property that $P(y > 0|\mathbf{x})$ and $E(y|y > 0, \mathbf{x})$ depend on different parameters so x_j can have dissimilar effects on these two functions (done in 'Advanced' Wooldridge).

Lecture 5 Outline

Poisson Regression Model

- A **count variable** is a dependent variable, which takes on nonnegative integer values $\{0, 1, 2, \dots\}$.
- Focus: when y takes on relatively few values including zero. Examples?
- Like Tobit outcome, can't take logarithm of a count variable because it takes on the value zero, so model expected value as an exponential function:

$$E(y|x_1, x_2, \dots, x_k) = \exp\left(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k\right) \tag{13}$$

Poisson Regression Model

$$\%\Delta E(y|\mathbf{x}) \approx (100\beta_j)\Delta x_j$$

So $100\beta_j$ is the approximate percentage change in $E(y|\mathbf{x})$ given a one-unit increase in x_j. When a more accurate estimate is needed, we can look at discrete changes in the expected value.

Keep all explanatory variables except x_k fixed and let x_k^0 be initial value and x_k^1 be subsequent value. Proportionate change in expected value is:

$$\left[\frac{\exp\left(\beta_0 + \mathbf{x_{k-1}}\boldsymbol{\beta}_{k-\mathbf{1}} + \beta_k x_k^1\right)}{\exp\left(\beta_0 + \mathbf{x_{k-1}}\boldsymbol{\beta}_{k-\mathbf{1}} + \beta_k x_k^0\right)}\right] - 1 = \exp\left(\beta_k \Delta x_k\right) - 1$$

We can interpret coefficients in model (13) as if we have a linear model with $\log(y)$ as the dependent variable.

Poisson Regression Model

- Since (13) is nonlinear in parameters ($\exp()$), we cannot use linear regression methods.
- We could use *nonlinear least squares* (NLLS) just like OLS minimizes SSR.
- However, all standard count data distributions are heteroscedastic and NLLS doesn't exploit this so we rely on ML and an important related method of *quasi-maximum likelihood estimation*.
- A count variable can't have a Normal distribution because the Normal distribution is for continuous variables that can take on all (or a large range [approx]) values and if it takes on very few values, the distribution is very different from a Normal, so instead the nominal distribution for count data is the **Poisson distribution**.

Poisson Regression Model

Since we're interested in effect of explanatory variables on y, look at Poisson distribution conditional on $\mathbf{x}$. Poisson distribution is determined solely by its mean so only need to specify $E(y|\mathbf{x})$. Assume this has same form as (13); write in shorthand as $\exp(\mathbf{x}\boldsymbol{\beta})$. Probability that y equals value h conditional on $\mathbf{x}$:

$$P(y = h|\mathbf{x}) = \frac{1}{h!} e^{-e^{\mathbf{x}\boldsymbol{\beta}}} \left(e^{\mathbf{x}\boldsymbol{\beta}}\right)^h \quad h = 0, 1, \ldots$$

The distribution forms basis for the **Poisson regression model**: can find conditional probability for any values of explanatory variables; e.g. $P(y = 0|\mathbf{x}) = e^{-e^{\mathbf{x}\boldsymbol{\beta}}}$. After estimating β_j, we can plug them into the probability for various values of $\mathbf{x}$. Given a random sample $\{(\mathbf{x}_i, y_i) : i = 1, \ldots, n\}$, log-likelihood is:

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \ell_i(\boldsymbol{\beta}) = \sum_{i=1}^{n} \{y_i \mathbf{x}_i \boldsymbol{\beta} - \exp(\mathbf{x}_i \boldsymbol{\beta})\} \qquad (14)$$

Poisson MLEs are not obtained in closed form.

Poisson Regression Model

Can't directly compare magnitude of Poisson estimates of exp () function with OLS estimates of linear function; rough comparison possible (at least for continuous explanatory variables).

While Poisson MLE is a natural first step for count data, it's often too restrictive. All probabilities and higher moments of Poisson distribution are determined entirely by mean. E.g:

$$Var(y|\mathbf{x}) = E(y|\mathbf{x}) \tag{15}$$

This has been shown to be violated in many applications.

Fortunately, the Poisson distribution has nice robustness property: whether or not Poisson distribution holds, we still get consistent, asymptotically Normal estimates of β_j like the Normality assumption for OLS (consistent and asymptotically Normal irrespective of the Normality assumption). Note: when we use Poisson MLE but not assume Poisson distribution is entirely correct, we call the analysis **quasi-maximum likelihood estimation (QMLE)**.

Poisson Regression Model

However, unless Poisson variance assumption (15) holds, SE need to be adjusted. Simple adjustment to SE is available when we assume variance is proportional to mean:

$$Var(y|\mathbf{x}) = \sigma^2 E(y|\mathbf{x}) \tag{16}$$

$\sigma^2 = 1 \Longrightarrow$ Poisson variance assumption. $\sigma^2 > 1 \Longrightarrow$ variance is greater than the mean for all $\mathbf{x}$ – called **overdispersion** since variance is larger than in the Poisson case and it is observed in many applications of count regressions. $\sigma^2 < 1 \Longrightarrow$ called *underdispersion* and is less common but is allowed in (16). Under (16), it's easy to adjust the usual Poisson MLE SE. Let $\hat{\beta}_j$ be Poisson QMLE and define residuals as $\hat{u}_i = y_i - \hat{y}_i$ where $\hat{y}_i = \exp\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \ldots + \hat{\beta}_k x_{ik}\right)$ is the fitted value. Residual for observation i is difference between y_i and fitted value. A consistent estimate of σ^2 is $\frac{1}{n-k-1}\sum_{i=1}^{n}\frac{\hat{u}_i^2}{\hat{y}_i}$

Poisson Regression Model

- Other count data models generalise Poisson.
- If we are only interested in effects of x_j on mean response, then there is little reason to go beyond Poisson: it's simple, often gives good results and has robustness property. We can actually apply Poisson to y that is a Tobit-like outcome, provided (13) holds, which might give good estimates of mean effects.
- Extensions of Poisson regression are more useful when interested in estimating probabilities, e.g. $P(y > 1|\mathbf{x})$ (see Cameron and Trivedi, 1998).

Lecture 5 Outline

Summary

- Corner solution models: Tobit.
  - Estimation by MLE through numerical methods.
  - Testing multiple hypotheses: Wald / LR.
  - Interpret estimates as partial effects and note inverse Mills ratio.
- Count models: Poisson.
  - Estimation by quasi-MLE.

References

- Tobit Model: Wooldridge 17.2.
- Poisson Regression Model: Wooldridge 17.3.