Econometrics

Week 11

Institute of Economic Studies
Faculty of Social Sciences
Charles University in Prague

Fall 2021

Recommended Reading

For today

- Limited Dependent Variable Models
- Chapter 17.2-17.5

For next week

- Repetition
- Chapters 10-17

Today's Talk

- Last week we learned how to deal with dummy dependent variables
  - Linear probability model (OLS)
  - Binary response model (e.g. Logit or Probit)

- This week we will discuss regression analysis with other types of limited dependent variables
  - Strictly positive dependent variable
  - Count variable
  - Censored variable
  - Truncated variable

Strictly positive dependent variable

- Let us consider a dependent variable that is (roughly) continuous over strictly positive values but zero otherwise.
- Such variables often appear in situations where corner solutions are common
  (e.g. money spent in restaurant per month)
- This behavior can be described using a latent variable model:

The Tobit Model

$$y^* = \beta_0 + x\beta + u, \qquad u|x \sim N(0, \sigma^2)$$

$$y = \max(0, y^*)$$

variable y^* satisfies the classical linear model assumptions (normal, homoscedastic distribution with linear conditional mean).

Strictly positive dependent variable

- To estimate parameters of the Tobit model, we use MLE
- The first step is to write down the likelihood function
  - Probability that $y = 0$ is $Prob(y^* \leq 0) = 1 - \Phi(x_i\beta/\sigma)$
  - Probability that $y > 0$ is $Prob(y^* > 0) = \Phi(x_i\beta/\sigma)$
  - Density for $y > 0$ is $(1/\sigma)\phi[(y_i - x_i\beta)/\sigma]$

MLE of the Tobit model

To estimate the parameters of the Tobit model, we need to maximize log-likelihood obtained by summing the following function across all observations i:

$$\begin{aligned} \ell_i(\beta,\sigma) &= \mathbb{1}(y_i = 0)\log[1 - \Phi(x_i\beta/\sigma)] \\ &\quad + \mathbb{1}(y_i > 0)\log\left((1/\sigma)\phi[(y_i - x_i\beta)/\sigma]\right), \end{aligned}$$

where $\mathbb{1}$ is an indicator function, Φ refers to standard normal cumulative distribution function, and ϕ to standard normal probability density function.

Interpreting the Tobit Model

- Any statistical software can be used to find parameters (β's and σ) that maximize this likelihood function. These parameters are called Tobit estimates.
- Like in case of logit/probit, also Tobit estimates are difficult to interpret
- According to the latent variable model, β's are effects of x on y^*. But we are interested in explaining y.
- We can analyze two values:
  - $E(y|y > 0, x) = x\beta + \sigma\lambda(x\beta/\sigma)$, see derivations on board where $\lambda(c) = \phi(c)/\Phi(c)$ is called the **Inverse Mills ratio**.
  - $E(y|x) = P(y > 0|x)E(y|y > 0, x) = \Phi(x\beta/\sigma)E(y|y > 0, x) = \Phi(x\beta/\sigma)\dot{x}\beta + \sigma\phi(x\beta/\sigma)$

  value of y conditional on $y > 0$ is equal to $x\beta$ plus a strictly positive term.

Interpreting the Tobit Model

- Usually we are interested in partial effects of x on y
- For continuous explanatory variable x_j:
  - $\frac{\partial E(y|y>0,x)}{\partial x_j} = \beta_j\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}$
    The adjustment factor is strictly between 0 and 1
  - $\frac{\partial E(y|x)}{\partial x_j} = \beta_j \Phi(x\beta/\sigma)$
  - Derive these formulas at home
- For discrete explanatory variable we calculate the difference in $E(y|y > 0, x)$ or $E(y|x)$ evaluated for $x_j = 1$ and $x_j = 0$
- Partial effects depend on β_j but also on σ and $x\beta$ (in the adjustment factor)

Interpreting the Tobit Model

- Usually we are interested in partial effects of x on y
- For continuous explanatory variable x_j:
  - $\frac{\partial E(y|y>0,x)}{\partial x_j} = \beta_j\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}$
    The adjustment factor is strictly between 0 and 1
  - $\frac{\partial E(y|x)}{\partial x_j} = \beta_j \Phi(x\beta/\sigma)$
  - Derive these formulas at home
- For discrete explanatory variable we calculate the difference in $E(y|y>0,x)$ or $E(y|x)$ evaluated for $x_j = 1$ and $x_j = 0$
- Partial effects depend on β_j but also on σ and $x\beta$ (in the adjustment factor)

Interpreting the Tobit Model

- Usually we are interested in partial effects of x on y
- For continuous explanatory variable x_j:
  - $\frac{\partial E(y|y>0,x)}{\partial x_j} = \beta_j\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}$
    The adjustment factor is strictly between 0 and 1
  - $\frac{\partial E(y|x)}{\partial x_j} = \beta_j \Phi(x\beta/\sigma)$
  - Derive these formulas at home
- For discrete explanatory variable we calculate the difference in $E(y|y > 0, x)$ or $E(y|x)$ evaluated for $x_j = 1$ and $x_j = 0$
- Partial effects depend on β_j but also on σ and $x\beta$ (in the adjustment factor)

Interpreting the Tobit Model

- Usually we are interested in partial effects of x on y
- For continuous explanatory variable x_j:
  - $\frac{\partial E(y|y>0,x)}{\partial x_j} = \beta_j\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}$
    The adjustment factor is strictly between 0 and 1
  - $\frac{\partial E(y|x)}{\partial x_j} = \beta_j \Phi(x\beta/\sigma)$
  - Derive these formulas at home
- For discrete explanatory variable we calculate the difference in $E(y|y>0,x)$ or $E(y|x)$ evaluated for $x_j = 1$ and $x_j = 0$
- Partial effects depend on β_j but also on σ and $x\beta$ (in the adjustment factor)

Interpreting the Tobit Model

- Usually we are interested in partial effects of x on y
- For continuous explanatory variable x_j:
  - $\frac{\partial E(y|y>0,x)}{\partial x_j} = \beta_j\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}$
    The adjustment factor is strictly between 0 and 1
  - $\frac{\partial E(y|x)}{\partial x_j} = \beta_j \Phi(x\beta/\sigma)$
  - Derive these formulas at home
- For discrete explanatory variable we calculate the difference in $E(y|y > 0, x)$ or $E(y|x)$ evaluated for $x_j = 1$ and $x_j = 0$
- Partial effects depend on β_j but also on σ and $x\beta$ (in the adjustment factor)

Interpreting the Tobit Model

- Usually we are interested in partial effects of x on y
- For continuous explanatory variable x_j:
  - $\frac{\partial E(y|y>0,x)}{\partial x_j} = \beta_j\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}$
    The adjustment factor is strictly between 0 and 1
  - $\frac{\partial E(y|x)}{\partial x_j} = \beta_j \Phi(x\beta/\sigma)$
  - Derive these formulas at home
- For discrete explanatory variable we calculate the difference in $E(y|y>0,x)$ or $E(y|x)$ evaluated for $x_j = 1$ and $x_j = 0$
- Partial effects depend on β_j but also on σ and $x\beta$ (in the adjustment factor)

Interpreting the Tobit Model

- Usually we are interested in partial effects of x on y
- For continuous explanatory variable x_j:
  - $\frac{\partial E(y|y>0,x)}{\partial x_j} = \beta_j\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}$
    The adjustment factor is strictly between 0 and 1
  - $\frac{\partial E(y|x)}{\partial x_j} = \beta_j \Phi(x\beta/\sigma)$
  - Derive these formulas at home
- For discrete explanatory variable we calculate the difference in $E(y|y > 0, x)$ or $E(y|x)$ evaluated for $x_j = 1$ and $x_j = 0$
- Partial effects depend on β_j but also on σ and $x\beta$ (in the adjustment factor)

Interpreting the Tobit Model

- Usually we are interested in partial effects of x on y
- For continuous explanatory variable x_j:
  - $\frac{\partial E(y|y>0,x)}{\partial x_j} = \beta_j\{1 - \lambda(x\beta/\sigma)[x\beta/\sigma + \lambda(x\beta/\sigma)]\}$
    The adjustment factor is strictly between 0 and 1
  - $\frac{\partial E(y|x)}{\partial x_j} = \beta_j \Phi(x\beta/\sigma)$
  - Derive these formulas at home
- For discrete explanatory variable we calculate the difference in $E(y|y>0,x)$ or $E(y|x)$ evaluated for $x_j = 1$ and $x_j = 0$
- Partial effects depend on β_j but also on σ and $x\beta$ (in the adjustment factor)
- The adjustment factor can be computed in two ways
  - At interesting values of x (usually the averages) - we get partial effect at the average (PEA)
  - For each observation separately and average afterwards - we get average partial effect (APE)

The Tobit Model - Discussion

- The Tobit model relays heavily on normality and homoskedasticity in the underlying latent variable.
  - Construction of the log-likelihood assumes normal, constant distribution
- In OLS, we can deal with heteroskedasticity for example by computing heteroskedasticity robust standard errors.
- But in Tobit, in case of heteroskedasticity we never know what MLE is actually estimating.

Count Dependent Variable

- **Count variable** takes only nonnegative integer values: $y_i = 0, 1, 2, ...$
- For example: number of children one has, number of car accidents, number of crimes committed...
- We want to model the distribution of y given the observed characteristics x.
  Which distribution function can be useful here?
- Let us assume that $y|x$ has the Poisson distribution
  - Let's assume that the expected value of y (e.g. the average number of children per woman with characteristics x) is: $E(y|x) = e^{x\beta}$
  - This assures positive values for $E(y|x)$
  - Then $Prob(y = h|x) = \frac{e^{(x\beta)^h} e^{-exp(x\beta)}}{h!}$
- The log-likelihood function: see board for derivations
  $\mathcal{L}(\beta) = \sum \ell_i(\beta) = \sum [y_i x_i \beta - exp(x_i \beta)]$

Count Dependent Variable

- **Count variable** takes only nonnegative integer values: $y_i = 0, 1, 2, ...$
- For example: number of children one has, number of car accidents, number of crimes committed...
- We want to model the distribution of y given the observed characteristics x.
  Which distribution function can be useful here?
- Let us assume that $y|x$ has the Poisson distribution
  - Let's assume that the expected value of y (e.g. the average number of children per woman with characteristics x) is: $E(y|x) = e^{x\beta}$
  - This assures positive values for $E(y|x)$
  - Then $Prob(y = h|x) = \frac{e^{(x\beta)^h} e^{-exp(x\beta)}}{h!}$
- The log-likelihood function: see board for derivations
  $\mathcal{L}(\beta) = \sum \ell_i(\beta) = \sum [y_i x_i \beta - exp(x_i\beta)]$

Count Dependent Variable

- **Count variable** takes only nonnegative integer values: $y_i = 0, 1, 2, ...$
- For example: number of children one has, number of car accidents, number of crimes committed...
- We want to model the distribution of y given the observed characteristics x.
  Which distribution function can be useful here?
- Let us assume that $y|x$ has the Poisson distribution
  - Let's assume that the expected value of y (e.g. the average number of children per woman with characteristics x) is: $E(y|x) = e^{x\beta}$
  - This assures positive values for $E(y|x)$
  - Then $Prob(y = h|x) = \frac{e^{(x\beta)^h} e^{-exp(x\beta)}}{h!}$

- The log-likelihood function: see board for derivations
  $\mathcal{L}(\beta) = \sum \ell_i(\beta) = \sum [y_i x_i \beta - exp(x_i\beta)]$

Count Dependent Variable

- **Count variable** takes only nonnegative integer values: $y_i = 0, 1, 2, ...$
- For example: number of children one has, number of car accidents, number of crimes committed...
- We want to model the distribution of y given the observed characteristics x.
  Which distribution function can be useful here?
- Let us assume that $y|x$ has the Poisson distribution
  - Let's assume that the expected value of y (e.g. the average number of children per woman with characteristics x) is: $E(y|x) = e^{x\beta}$
  - This assures positive values for $E(y|x)$
  - Then $Prob(y = h|x) = \frac{e^{(x\beta)^h} e^{-exp(x\beta)}}{h!}$
- The log-likelihood function: see board for derivations
  $\mathcal{L}(\beta) = \sum \ell_i(\beta) = \sum [y_i x_i \beta - exp(x_i\beta)]$

Count Dependent Variable

- **Count variable** takes only nonnegative integer values: $y_i = 0, 1, 2, ...$
- For example: number of children one has, number of car accidents, number of crimes committed...
- We want to model the distribution of y given the observed characteristics x.
  Which distribution function can be useful here?
- Let us assume that $y|x$ has the Poisson distribution
  - Let's assume that the expected value of y (e.g. the average number of children per woman with characteristics x) is: $E(y|x) = e^{x\beta}$
  - This assures positive values for $E(y|x)$
  - Then $Prob(y = h|x) = \frac{e^{(x\beta)^h} e^{-exp(x\beta)}}{h!}$
- The log-likelihood function: see board for derivations
  $\mathcal{L}(\beta) = \sum \ell_i(\beta) = \sum [y_i x_i \beta - exp(x_i \beta)]$

Count Dependent Variable

- **Count variable** takes only nonnegative integer values: $y_i = 0, 1, 2, ...$
- For example: number of children one has, number of car accidents, number of crimes committed...
- We want to model the distribution of y given the observed characteristics x.
  Which distribution function can be useful here?
- Let us assume that $y|x$ has the Poisson distribution
  - Let's assume that the expected value of y (e.g. the average number of children per woman with characteristics x) is: $E(y|x) = e^{x\beta}$
  - This assures positive values for $E(y|x)$
  - Then $Prob(y = h|x) = \frac{e^{(x\beta)^h} e^{-exp(x\beta)}}{h!}$
- The log-likelihood function: see board for derivations
  $\mathcal{L}(\beta) = \sum \ell_i(\beta) = \sum [y_i x_i \beta - exp(x_i \beta)]$

Count Dependent Variable

- **Count variable** takes only nonnegative integer values: $y_i = 0, 1, 2, ...$
- For example: number of children one has, number of car accidents, number of crimes committed...
- We want to model the distribution of y given the observed characteristics x.
  Which distribution function can be useful here?
- Let us assume that $y|x$ has the Poisson distribution
  - Let's assume that the expected value of y (e.g. the average number of children per woman with characteristics x) is: $E(y|x) = e^{x\beta}$
  - This assures positive values for $E(y|x)$
  - Then $Prob(y = h|x) = \frac{e^{(x\beta)^h} e^{-exp(x\beta)}}{h!}$
- The log-likelihood function: see board for derivations
  $\mathcal{L}(\beta) = \sum \ell_i(\beta) = \sum [y_i x_i \beta - exp(x_i \beta)]$

Count Dependent Variable

- **Count variable** takes only nonnegative integer values: $y_i = 0, 1, 2, ...$
- For example: number of children one has, number of car accidents, number of crimes committed...
- We want to model the distribution of y given the observed characteristics x.
  Which distribution function can be useful here?
- Let us assume that $y|x$ has the Poisson distribution
  - Let's assume that the expected value of y (e.g. the average number of children per woman with characteristics x) is: $E(y|x) = e^{x\beta}$
  - This assures positive values for $E(y|x)$
  - Then $Prob(y = h|x) = \frac{e^{(x\beta)^h} e^{-exp(x\beta)}}{h!}$
- To estimate parameters of the Poisson function we turn to MLE again
- The log-likelihood function: see board for derivations
  $\mathcal{L}(\beta) = \sum \ell_i(\beta) = \sum [y_i x_i\beta - exp(x_i\beta)]$

Interpreting Poisson Regression

- Any statistical software can be used to maximize the likelihood function and obtain Poisson estimates
- Partial effect of x_j on y:

$$\frac{\partial E(y|x)}{\partial x_j} = e^{x\beta}\beta_j$$

- The APE scaling factor is: $\frac{\sum exp(x\widehat{\beta})}{n} = \frac{\sum \hat{y}_i}{n} = \bar{y}$
- Note that Poisson ML estimator is consistent and asymptotically normal even if the data don't follow Poisson distribution!
  - In this case we use Quasi-Maximum Likelihood estimation

Truncated Regression Models

- A **truncated regression model** arises when we exclude a subset of population from the sampling scheme.
- In other words, we do not have random sampling, but have some rule to sample the data.
- For example, if it is not possible to survey part of population for any reason.

The Truncated Normal Regression Model

$$y = \beta_0 + x\beta + u, \qquad u|x \sim N(0, \sigma^2)$$

model would satisfy all but one CLM assumptions. We do not have a random sampleOLS suffers from sample selection bias!

Sample Selection Correction

- Truncated regression is a special case of general problem: **nonrandom sample selection**. can also think about it as omitted variable problem: we miss the information on how the sample was chosen. way how to correct for it and obtain consistent estimates of β is to use: $E(y|z, s=1) = x\beta + \rho\lambda(z\gamma)$, z is set of instrumental variables: $z\gamma = \gamma_0 + \gamma_1 z_1 + \ldots + \gamma_m z_m$.

Sample Selection Correction

all observations, estimate a probit model of selection indicator s_i on z_i. the inverse Mills ratio $\hat{\lambda}_i = \lambda(z_i\hat{\gamma})$. the selected sample (observations where $s_i = 1$), run regression of y_i on x_i and $\hat{\lambda}_i$.

Sample Selection Correction

- We can use a simple test for selection bias using this procedure. can test $H_0 : \rho = 0$ under which there is no sample selection problem.
- Against $H_A : \rho \neq 0$ under which there is sample selection problem t statistic on $\hat{\lambda}_i$. $\rho \neq 0$, OLS errors are not correct.

Censored Regression Model

- **Censored regression model** has a similar structure to the Tobit model.
- These two are sometimes interchanged, but they are different.
  - In Tobit the negative values of y do not exist
  - In censored regression the values below/above a threshold exist, but are **censored**
- We assume the underlying dependent variable to be normally distributed and censored below or above a certain value due to the way we collect data.
  - E.g. Earnings above 100 000 Kc/month are reported without stating the exact amount

Censored Regression Models

- Generally, censored regression model can be defined without distributional assumptions. will study them under assumption of normal distribution.

The Censored Normal Regression Model

$$\begin{aligned} y_i &= \beta_0 + x_i\beta + u_i, \qquad u_i|x_i, c_i \sim N(0, \sigma^2) \\ w_i &= \min(y_i, c_i), \end{aligned}$$

with censoring value c_i.

we only observe $w_i = \min(y_i, c_i)$ in right censored data, left censored data, we observe $w_i = \max(y_i, c_i)$.

Censored Regression Models

- A good example of censored data is some type of surveys. we ask respondents about their wealth, but they are allowed to respond with "more than 1.000.000 CZK", then we observe actual wealth of people below this threshold, but not above. threshold c_i in this case is constant for all i. many situations, censoring threshold changes with individuals i. are able to estimate the censored regression model using MLE.

Thank you

Thank you very much for your attention!