Introductory Econometrics

Lecture 25: Limited dependent variable models

Instructor: Ma, Jun

Renmin University of China

June 7, 2023

Data censoring

- The explained variable of interest may not be perfectly observed.
- The explained variable can be censored. e.g. Income data are often top-coded in survey data. The annual incomes above 200000 may be loaded as 200000. Households with higher incomes than 200000 are part of the sample and their characteristics are reported.

The Tobit model

- Consider a linear latent random variable Y^* (e.g. the real income) can be explained by a linear model in X (assuming there is a single explanatory variable for simplicity):

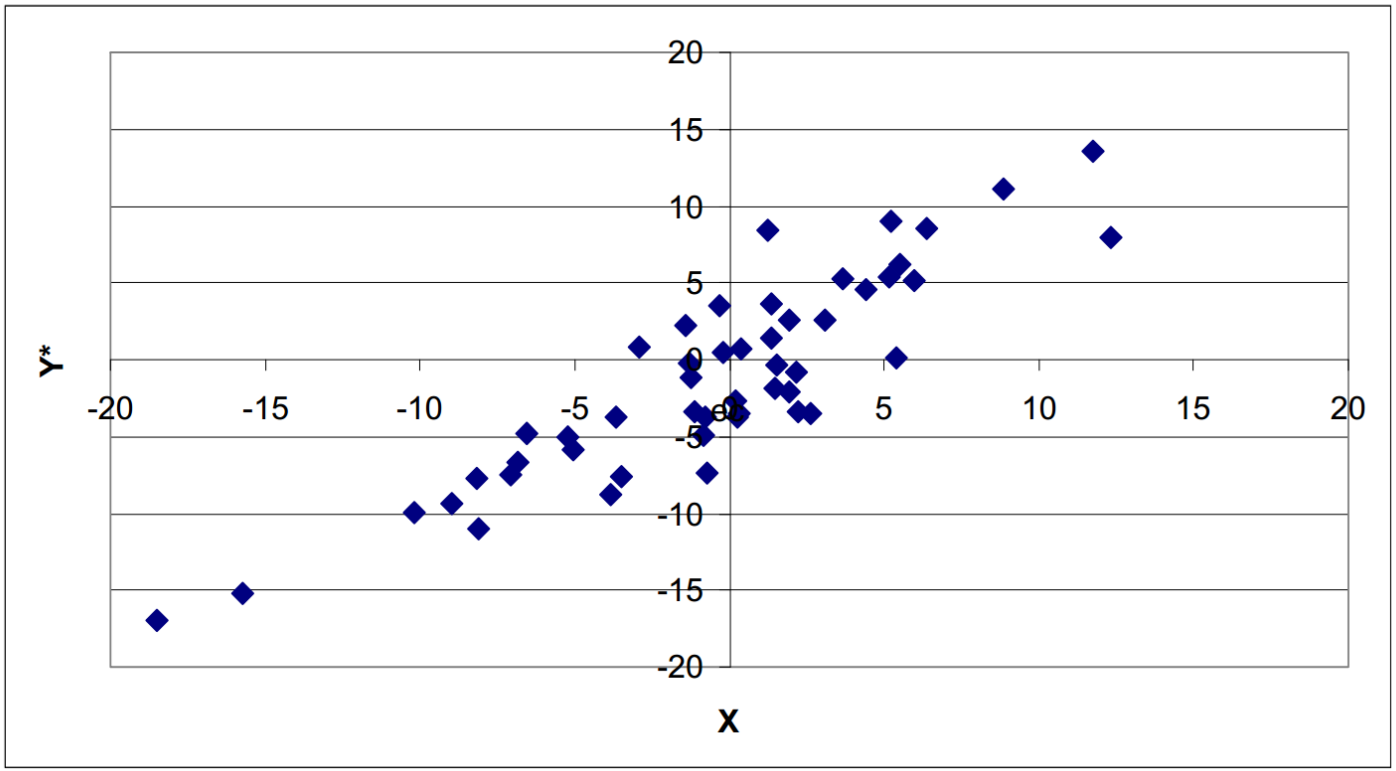
$$Y_i^* = \alpha + \beta X_i + \epsilon_i$$

  where ϵ is independent from X and distributed as $\mathrm{N}\left(0, \sigma^2\right)$.
- The distribution of Y^* conditionally on X is therefore normal: $Y^* \mid X \sim \mathrm{N}\left(\alpha + \beta X, \sigma^2\right)$.
- The latent model is assumed to be homoskedastic, since $\mathrm{E}\left[\epsilon^2 \mid X\right]$ is an unknown constant σ^2.
- The observed explained variable Y_i is censored below 0:

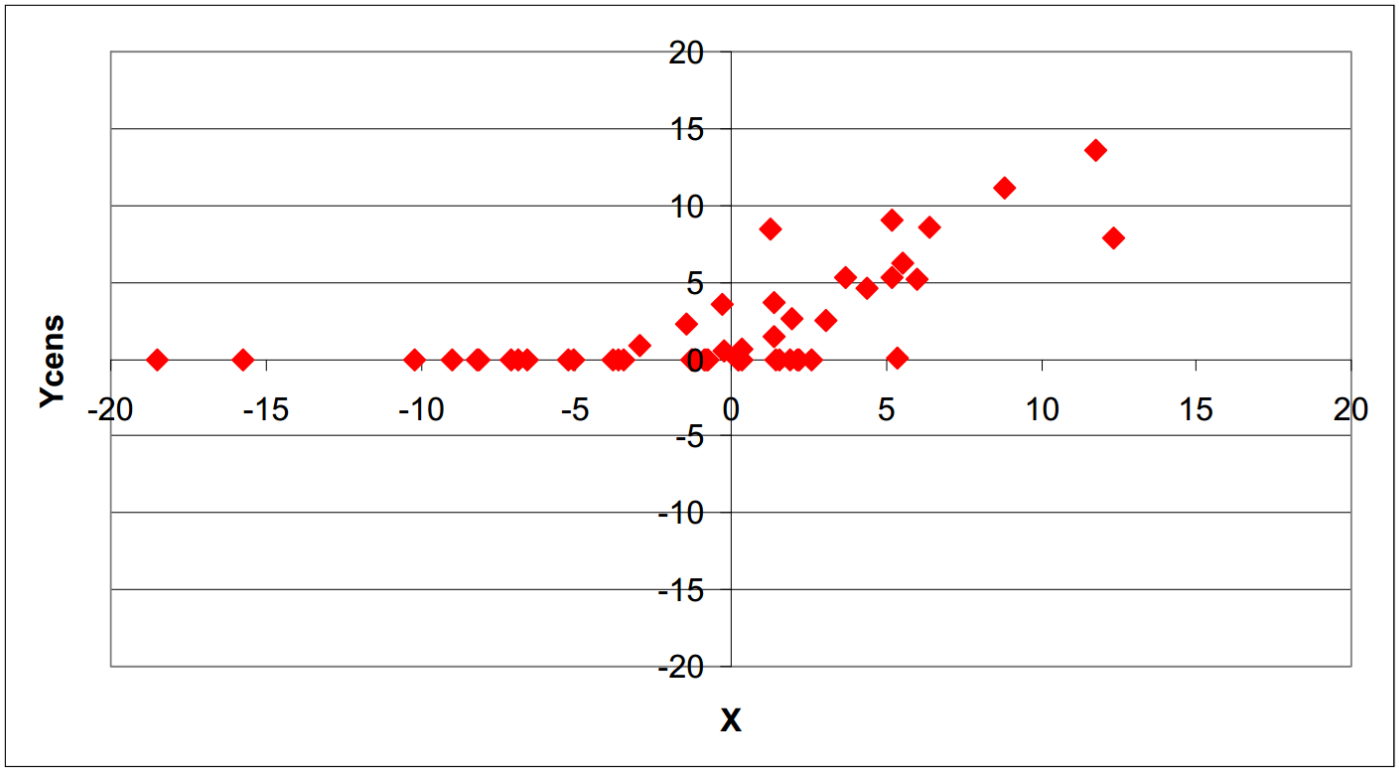
$$Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > 0 \\ 0 & \text{if } Y_i^* \leq 0, \end{cases}$$

  in other words, $Y_i = \max\left\{Y_i^*, 0\right\}$. This is called censoring from below at 0.

Uncensored data



Censored data



The conditional expectation of Y

- The conditional expectation of Y given X is

$$\mathrm{E}[Y \mid X] = \Pr[Y = 0 \mid X] \times 0 + \Pr[Y > 0 \mid X] \, \mathrm{E}[Y \mid Y > 0, X].$$

- By independence of ϵ,

$$\begin{aligned} \mathrm{P}[Y > 0 \mid X] &= \Pr[Y > 0 \mid X] \\ &= \Pr[\epsilon > -(\alpha + \beta X) \mid X] \\ &= \Phi\left(\frac{\alpha + \beta X}{\sigma}\right). \end{aligned}$$

- Let $Z \sim \mathrm{N}\left(\mu, \sigma^2\right)$, then we have

$$\mathrm{E}\left[Z \mid Z > d\right] = \mu + \sigma \frac{\phi\left(\frac{d-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{d-\mu}{\sigma}\right)}.$$

- This is a special property of normal distributions.
- Using this result, we can obtain

$$\begin{aligned}
\mathrm{E}\left[Y \mid Y > 0, X\right] &= \mathrm{E}\left[\alpha + \beta X + \epsilon \mid \epsilon > -\left(\alpha + \beta X\right), X\right] \\
&= \alpha + \beta X + \mathrm{E}\left[\epsilon \mid \epsilon > -\left(\alpha + \beta X\right), X\right] \\
&= \alpha + \beta X + \sigma \frac{\phi\left(\frac{\alpha+\beta X}{\sigma}\right)}{1 - \Phi\left(\frac{\alpha+\beta X}{\sigma}\right)}.
\end{aligned}$$

- Therefore we have

$$\mathrm{E}[Y \mid X] = \Phi\left(\frac{\alpha+\beta X}{\sigma}\right)\left(\alpha+\beta X+\sigma\frac{\phi\left(\frac{\alpha+\beta X}{\sigma}\right)}{\Phi\left(\frac{\alpha+\beta X}{\sigma}\right)}\right)$$

  which is nonlinear in the parameters.
- A linear regression of Y on X's yields an inconsistent estimator for β.
- $Y = \mathrm{E}[Y \mid X] + U$ and by definition, $\mathrm{E}[U \mid X] = 0$. In the fitted model

$$Y = \alpha + \beta X + V$$

  where $V = \mathrm{E}[Y \mid X] - (\alpha+\beta X) + U$ is in general correlated with X.

Maximum likelihood estimation

- The conditional distribution of Y given X is like the distribution of a mixture of a discrete random variable and a continuous random variable.
- The conditional density of Y given X is

$$f_{Y|X}(y \mid x, \alpha, \beta, \sigma) = \left[1 - \Phi\left(\frac{\alpha + \beta x}{\sigma}\right)\right]^{1[y=0]} \left[\frac{1}{\sigma}\phi\left(\frac{y - \alpha - \beta x}{\sigma}\right)\right]^{1[y>0]}$$

  for $y \geq 0$.
- The log-likelihood function is given by

$$\begin{aligned} \log L(a, b, c) &= \sum_{i=1}^{n} 1[Y_i = 0] \log\left(1 - \Phi\left(\frac{a + bX_i}{c}\right)\right) \\ &\quad + \sum_{i=1}^{n} 1[Y_i > 0] \log\left(\frac{1}{c}\phi\left(\frac{Y_i - a - bX_i}{c}\right)\right). \end{aligned}$$

- The maximizer of the log-likelihood function with respect to (a, b, c) is the maximum likelihood estimator for $\left(\alpha, \beta, \sigma^2\right)$.

Marginal effect

- It can be shown that the marginal effect of the explanatory variable X on the censored explained variable Y is

$$\frac{\mathrm{dE}\left[Y \mid X = x\right]}{\mathrm{d}x} = \beta\Phi\left(\frac{\alpha + \beta x}{\sigma}\right).$$

- The estimated average marginal effect is

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\beta}\Phi\left(\frac{\hat{\alpha} + \hat{\beta}X_i}{\hat{\sigma}}\right)$$

by plugging in the maximum likelihood estimates $\hat{\beta}$, $\hat{\sigma}$.

Stata implementation

```
. tobit mpg wgt, ul(24)
Tobit regression                                Number of obs     =         74
                                                LR chi2(1)        =      90.72
                                                Prob > chi2       =     0.0000
Log likelihood =  -129.8279                     Pseudo R2         =     0.2589

------------------------------------------------------------------------------
         mpg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         wgt |  -5.080645     .43493   -11.68   0.000    -5.947459   -4.213831
       _cons |   36.08037   1.432056    25.19   0.000     33.22628    38.93445
-------------+----------------------------------------------------------------
       /sigma |   2.385357   .2444604                      1.898148    2.872566
------------------------------------------------------------------------------

  Obs. summary:          0  left-censored observations
                        51     uncensored observations
                        23 right-censored observations at mpg>=24
```

In stata we can estimate more general models with censoring from above (option: ul(#)) and below (option: ll(#)).

Truncated samples

- Data on dependent and independent variables sampled from a sub-population, based on value of dependent variable.
- For truncated samples, data is simply not available to the researcher.

The truncated regression model

- Consider a linear latent random variable Y^* can be explained by a linear model in X (assuming there is a single explanatory variable for simplicity):

$$Y_i^* = \alpha + \beta X_i + \epsilon_i$$

  where ϵ is independent from X and distributed as $\mathrm{N}\left(0, \sigma^2\right)$.
- The distribution of Y^* conditionally on X is therefore normal: $Y^* \mid X \sim \mathrm{N}\left(\alpha + \beta X, \sigma^2\right)$.
- The observed explained variable Y_i is truncated below 0:

$$Y_i = \begin{cases} Y_i^* & \text{if } Y_i^* > 0 \\ \text{not available} & \text{if } Y_i^* \leq 0. \end{cases}$$

Conditional density

- The conditional density for a random variable U (given X, we need only to consider the distribution of the error term) with unconditional density f, cumulative distribution function F, and truncation below at c is

$$f_U(u \mid U > c) = \begin{cases} \frac{f(u)}{\mathrm{P}[U>c]} = \frac{f(u)}{1-F(c)} & \text{if } u > c \\ 0 & \text{otherwise} \end{cases}.$$

- This is still a density. It is clearly positive and integration gives

$$\int_{-\infty}^{\infty} f_U(u \mid U > c)\,\mathrm{d}u = \int_c^{\infty} \frac{f(u)}{1-F(c)}\mathrm{d}u = \frac{1-F(c)}{1-F(c)} = 1.$$

Maximum likelihood estimation

- The conditional distribution of the observed truncated variable Y given X is like the distribution of a mixture of a discrete random variable and a continuous random variable.
- The conditional density of Y given X is

$$f_{Y|X}(y \mid x, \alpha, \beta, \sigma) = \frac{f_{Y^*|X}(y \mid x)}{\mathrm{P}[Y^* > 0 \mid X = x]} = \frac{\sigma^{-1}\phi\left(\frac{y-\alpha-\beta x}{\sigma}\right)}{1-\Phi\left(\frac{-\alpha-\beta x}{\sigma}\right)}, \; y \geq 0.$$

- The log-likelihood function is given by

$$\log L(a, b, c) = \sum_{i=1}^{n} \log\left(\frac{1}{c}\phi\left(\frac{Y_i - a - bX_i}{c}\right)\right) - \sum_{i=1}^{n} \log\left(1-\Phi\left(\frac{-a-bX_i}{c}\right)\right).$$

- The maximizer of the log-likelihood function with respect to (a, b, c) is the maximum likelihood estimator for (α, β, σ).

Sample selection

- Sample selection problem occurs when the observed sample is not a random sample but systematically chosen from the population.
- The classical example: we want to explain the market wage of married women, but a large fraction of the respondents decided to stay at home.

The selection model

- Consider a model with two latent variables Y_i^* and D_i^* ($i = 1, \ldots, n$) which linearly depend on observable explanatory variables X_i, Z_i:

$$D_i^* = \gamma Z_i + V_i \text{ (participation equation)}$$
$$Y_i^* = \beta X_i + U_i \text{ (outcome equation)}$$

with

$$(V_i, U_i) \sim \mathrm{N}\left(0, \begin{bmatrix} 1 & \sigma_{uv} \\ \sigma_{uv} & \sigma_u^2 \end{bmatrix}\right).$$

The error terms U_i and V_i are independently and jointly normally distributed with covariance σ_{uv}.

- The two latent variables cannot be observed by the researcher.
- We only observe an indicator when the latent variable D_i^* is positive:

$$D_i = \begin{cases} 1 & \text{if } D_i^* > 0 \\ 0 & \text{otherwise} \end{cases}.$$

- The value of the variable Y_i is only observed if the indicator is 1:

$$Y_i = \begin{cases} Y_i^* & \text{if } D_i = 1 \\ \text{not available} & \text{otherwise} \end{cases}.$$

- The first equation explains whether an observation is in the sample or not.
- The second equation determines the value of Y_i.
- For explaining the market wage of married women, Y_i^* is the market wage of individual i, i.e. the wage she would have if participating work.
- D_i^* could be a latent index that can be thought of as representing the difference between the observed wage and the reservation wage, the lowest wage at which the individual is willing to participate work.
- In real applications, Z_i and X_i could be vectors of different dimensions and contain different explanatory variables.

The conditional expectation

- The expected value of the observed variable Y_i conditional on it being observed is

$$\mathrm{E}\left[Y_i \mid D_i = 1, X_i, Z_i\right] = \beta X_i + \sigma_{uv} \frac{\phi\left(\gamma Z_i\right)}{\Phi\left(\gamma Z_i\right)} = \beta X_i + \sigma_{uv} \lambda\left(\gamma Z_i\right)$$

  where $\lambda(\cdot) = \phi(\cdot) / \Phi(\cdot)$ is called the inverse Mills ratio.
- A linear regression of observed Y_i's on X_i is an inconsistent estimator for β.
- If the errors are uncorrelated (i.e. $\sigma_{uv} = 0$), then a simple linear regression gives a consistent estimator.
- If γ is known, then we can construct a regressor $\lambda\left(\gamma Z_i\right)$ and regress observed Y_i on X_i and $\lambda\left(\gamma Z_i\right)$. This procedure gives a consistent estimator for β.
- Two methods to consistently estimate β (the parameter of interest) is (1). Maximum Likelihood (2). Heckman two-step procedure.

Heckman two-step procedure

- In practice, γ is unknown. Since we observe D_i and Z_i, we can consistently estimate γ using a Probit estimator.
- This leads to a two-step procedure.
- In step 1, we estimate a Probit model and obtain the Probit estimator $\hat{\gamma}$ for γ.
- In step 2, we construct $\lambda(\hat{\gamma}Z_i)$ and linearly regress Y_i on X_i and $\lambda(\hat{\gamma}Z_i)$. The estimate of the slope corresponding to X_i is a consistent estimator for β.
- The Heckman estimator for β is asymptotically normal. However, if using STATA to explicitly compute the two-step estimator, the standard errors in the second-step output use the wrong formula.
- Standard errors must be corrected because in the second step, the regressor $\lambda(\hat{\gamma}Z_i)$ is generated from the first-step Probit model.