ECON 4151
Lab Session 6: Poisson model and selectivity model

Wonjin Lee

Washington University in St. Louis

November 20, 2020

Outline

1. Poisson regression model (Wooldridge Ch. 17)

2. Selectivity models (Wooldridge Ch. 17)

Poisson regression model

- Poisson distribution:

$$\Pr(y = h|\lambda) = \frac{e^{-\lambda}\lambda^h}{h!}, \quad h = 0, 1, \ldots$$

  with $\mathbb{E}[y] = \lambda$ and $\text{Var}[y] = \lambda$.
  - The poisson distribution is entirely determined by its mean, λ.
- In the Poisson regression model, assume that

$$\lambda = \mathbb{E}[y|x] = e^{x'\beta}$$

  - $\log(\mathbb{E}[y|x]) = x'\beta$
  - We cannot take the logarithm of a count variable because it takes on the value zero; model the expected value as an exponential function.
- Then, its distribution is

$$\Pr(y = h|x) = \frac{e^{-e^{x'\beta}}(e^{x'\beta})^h}{h!}, \quad h = 0, 1, \ldots$$

- Partial effect at the average and average partial effect are defined just like logit/probit models.

Why Poisson distribution?

- A count variable cannot have a normal distribution b/c the normal distribution is for continuous variables that can take on all values.
- Also, if it takes on very few values, the distribution can be very different from normal.
- Therefore, Poisson distribution is used for count data.

MLE

- Log-likelihood function:

$$\ln \mathcal{L}(\beta) = \sum_{i=1}^{N} \left[y_i x_i' \beta - e^{x_i' \beta} \right]$$

- Maximize it to get $\hat{\beta}$.

QMLE

- Quasi-maximum likelihood estimation (QMLE): We use Poisson MLE but we do not assume that the Poisson distribution is entirely correct.
- A simple adjustment to the standard errors:

$$\mathsf{Var}[y|x] = \sigma^2 \mathbb{E}(y|x)$$

where a consistent estimator of $\sigma^2 = \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{(n-k-1)\hat{y}_i}$.

Exercise

- Wooldridge Example 17.3: Determinants of number of arrests for young men

Selectivity: Censored regression model

- Censored normal regression model (generalized Tobit model):

$$y_i^* = x_i'\beta + u_i, \quad u|x, c \sim \mathcal{N}\left(0, \sigma^2\right),$$
$$y_i = \min\left\{c_i, y_i^*\right\}$$

  - This is censored from above, or right censored.
  - We only know that the variable is at least as large as the threshold.
  - Ex. top coding (its value is known only up to a certain threshold)
- The density of y given x and c:
  - For $y < c_i$

$$f\left(y_i|x_i, c_i\right) = \frac{1}{\sigma}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)$$

  - For $y = c_i$

$$P\left(y_i = c_i|x_i\right) = 1 - \Phi\left(\frac{c_i - x_i'\beta}{\sigma}\right)$$

- Maximize the log-likelihood function to get $\hat{\beta}$ and $\hat{\sigma}$.

Application of censored normal regression model

- Duration analysis, top coding, ...
  - Duration: a variable that measures the time before a certain event occurs.
- We often use the natural log as the dependent variable, which means we also take the log of the censoring threshold.
  1. The parameters are interpreted as percentage changes.
  2. The log of a duration typically has a distribution closer to (conditional) normal than the duration itself.

Selectivity: Truncated regression model

- Truncated normal regression model:

$$y_i = x_i'\beta + u_i, \quad u|x \sim \mathcal{N}\left(0, \sigma^2\right)$$

- Data truncation: (x_i, y_i) is observed only if $y_i \leq c_i$.
    - We choose the sample on the basis of the dependent variable.
    - We do not observe a random sample from the population; a subset of the population prior to sampling.
- The density of y given x and c:

$$g\left(y_i|x_i, c_i\right) = \frac{f\left(y_i|x_i'\beta, \sigma^2\right)}{F\left(c_i|x_i'\beta, \sigma^2\right)}, \quad y_i \leq c_i$$

- Maximize the log-likelihood function to get $\hat{\beta}$ and $\hat{\sigma}$.

Final remark on selectivity model

- Interpret the β just as in a linear regression model under random sampling.
- This is much different than Tobit applications to corner solution responses, where the expectations of interest are nonlinear functions of the β
- Censored regression models may be preferable to the truncated regression models because in the former we include all the observations in the sample, whereas in the latter we only include observations in the truncated sample.

Exercise

- Wooldridge Example 17.4: Duration of recidivism
- Gujarati 11.3: Married women's annual labor supply