Limited Dependent Variable Models

Semester 2/2013

Part 1

Chayanee Chawanote

Limited dependent variable

- Limited dependent variable (LDV): a dependent variable whose range of values is substantively restricted. It can be a continuous variable or a discrete choice. For example,
  - A binary variable has only 2 values, zero and one.
  - Transportation choice: bus, subway, private car, taxi
  - Health status: poor, fair, good, excellent
  - Income tax data recorded only if individuals have to pay an income tax

A Binary dependent variable

The Linear Probability Model (LPM)

- Let y = 1 be one of the outcomes that we are interested, and y = 0 the other outcome.
- Suppose the population model is
  $y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$
  - How can we interprete β when y is either 0 or 1?
- Think about 'success' and 'failure' as in Bernoulli distribution.
- P(y=1|x) = E(y|x): the probability of 'success' is a linear function of the x_j
  $P(y = 1|x) = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k$ (1)
- Call equation (1) the linear probability model

Linear probability model

- The response probability, P(y=1|x), is linear in the parameter β_j
- β_j measures the change in the probability of success when x_j changes, holding other factors fixed:
  $\triangle P(y = 1|x) = \beta_j \triangle x_j$
- If the estimated equations is $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k$
- $\hat{y}$ is the predicted probability of success.
- $\hat{\beta}_j$ measures the predicted change in the probability of success when x_j increases by one unit

Linear probability model

- Shortcomings of the LPM:
  - It's likely to get predicted probability either less than 0, or greater than 1.
  - A probability cannot be linearly related to the independent variables for all their possible values.
  - There is heteroskedasticity in a linear probability model since variance of a binary variable is $Var(y|x) = p(x)[1 - p(x)]$, depending on x.
- We can calculate the "percent correctly predicted", instead of using the fitted values that might exceed 1
  - Define a predicted value as $\tilde{y}_i = 1$ if $\hat{y}_i < 0.5$ so that we have the proportion of overall correct predictions