Applied Microeconometrics (L7): Binary choice models

Nicholas Giannakopoulos

University of Patras
Department of Economics

ngias@upatras.gr

November 26, 2019

Overview

Modeling

Linear Probability Model

Logit

Probit

Modelling

- Continuous/quantitative variable. Examples:
  - Economic growth
  - Log of value-added or output
  - Log of earnings, etc.
- Not continuous/not quantitative variable. Examples:
  - What characteristics (e.g. parental) affect the likelihood that an individual obtains a higher degree? (1:Yes, 0:No)
  - What determines labour force participation? (1:Yes, 0:No)
  - Why unemployed search for job? (1:Yes, 0:No)
  - Why to buy a car? (1:Yes, 0:No)
  - Why a firm invests in new machinery (1:Yes, 0:No)
  - What factors drive the incidence of civil war? (1:Yes, 0:No)

Binary Response Models

- Basics: think in terms of probabilities
  - "What is the probability that an individual with certain characteristics owns a car?"
  - "If some variable X changes by one unit, what is the effect on the probability of owning a car?"
- When the dependent variable y is binary: 1 for observations in the dataset for which the event of interest has happened ("success"), 0 otherwise ("failure").
- In a random sample, the sample mean of y is an unbiased estimate of the unconditional probability that the event happens.
  - $Pr(y = 1) = E(y) = \frac{\sum_i^n y_i}{N}$
  - $i = 1, ..., n$, N: number of observations
- Unconditional probability: trivial
- Conditional probability: apply regression analysis

Binary Response Models: theoretical framework

$$y = \begin{cases} 1, & \text{with probability p} \\ 0, & \text{with probability 1-p} \end{cases}$$

- Binary dependent variable y : (0/1)
- Explanatory variables x: continuous and/or discrete
- Conditional expectation of y given x

$$E(y|x) = Pr(y = 1|x)$$

- Standard regression framework

$$y = F(y, \beta) + u$$

- Classical assumptions of regression analysis

$$E(y|x) = F(y, \beta) + E(u|x) = F(y, \beta)$$

- Specify the functional form of $F(y, \beta)$
  - Simplest functional form: Linear Probability Model

Linear Probability Model: Set Up

- Dependent variable: $y : (0, 1)$
- Explanatory variables: $k \times 1$ vector of x's
- Conditional probability: $Pr(y = 1|x) = F(x, \beta) = x'\beta$
- Sample: n observations (x_1, y_i) drawn randomly from a population
- Estimation method: OLS

$$y_i = x_i'\beta + u_i$$

- β measures the change in the probability of "success", resulting from a change in the variable x, holding other factors fixed (i.e., partial effect on the probability of "success")
- $\Delta Pr(y = 1|x) = \beta \Delta x$
- Example: use mus14data.dta

Problems with LPM

- Dependent variable: $y : (0, 1)$
- Explanatory variables: $k \times 1$ vector of x's
- Conditional probability: $Pr(y = 1|x) = F(x, \beta) = x'\beta$
- Sample: n observations (x_1, y_i) drawn randomly from a population
- Estimation method: OLS

$$y_i = x_i'\beta + u_i$$

- β measures the change in the probability of "success", resulting from a change in the variable x, holding other factors fixed (i.e., partial effect on the probability of "success")
- $\Delta Pr(y = 1|x) = \beta \Delta x$
- Example: use mus14data.dta

Logit model

$$Prob(y = 1) = \Lambda(x'\beta) = \frac{e^{x'\beta}}{1 + e^{x'\beta}} \quad (1)$$

where $\Lambda(\cdot)$ is the logistic cumulative density function (cdf) with $\Lambda(z) = \frac{e^z}{1+e^z} = \frac{1}{(1+e^{-z})}$. Use Maximum Likelihood Estimation (MLE) techniques to get estimates. Stata command: logit. For presentation purposes estimate Marginal Effects. Stata command: margins.

Probit model

$$Prob(y = 1) = \Phi(x'\beta) = \int_{-\infty}^{x'\beta} \phi(z)dz \qquad (2)$$

where $\Phi(\cdot)$ is the standard cumulative density function (cdf) with derivative $\phi(z) = \left(\frac{1}{\sqrt{2\pi}}\right) e^{\left(\frac{-z^2}{2}\right)}$ which is the standard normal density function. Use Maximum Likelihood Estimation (MLE) techniques to get estimates. Stata command: probit. For presentation purposes estimate Marginal Effects. Stata command: margins.