Econometrics

Chapter 15: Logit and Probit models

In Choi

Sogang University

Linear Probability Model

- When y_i is a binary variable (i.e., 0 or 1)

$$E(y_i|x) = 1 \times P(y_i = 1|x) + 0 \times P(y_i = 0|x) = P(y_i = 1|x),$$

so we can write our model as

$$P(y_i = 1|x) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$$

- As in the standard linear regression, regress y_i on $x's$ – easy to use.
- The interpretation of β_j is the change in the probability of success ($y_i = 1$) when x_j changes: $\frac{dp_i}{dxij} = \beta_j$.

Linear Probability Model

- The predicted y_i is the predicted probability of success. A potential problem that it can be outside [0,1].
- We may obtain slope estimates which imply that a change in x changes the probability by more than $+1$ or –1.

Logit and Probit Models

- Logit model assumes

$$P(y_i = 1|x) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}.$$

- $\frac{dP(y_i=1|x)}{dx_{ij}} = \frac{e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}}{\left(1 + e^{\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}}\right)^2} \beta_j$
- Interpreting the coefficient estimate is not straightforward.
- But the signs of $\frac{dP(y_i=1|x)}{dx_{ij}}$ and β_j are the same.

Logit and Probit Models

- Probit model assumes

$$\begin{aligned} P(y_i &= 1|x) = P(N(0,1) < \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}) \\ &= \int_{-\infty}^{\beta_0+\beta_1 x_{i1}+\ldots+\beta_k x_{ik}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt. \end{aligned}$$

- $\frac{dP(y_i=1|x)}{dx_{ij}} = \phi(\beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik})\beta_j$ (ϕ is the pdf of a standard normal variable.)
- Interpreting the coefficient estimate is not straightforward.
- But the signs of $\frac{dP(y_i=1|x)}{dx_{ij}}$ and β_j are the same.

Logit and Probit Models

- Probit and probit models are estimated by the maximum likelihood methods (cf. Wooldridge, chapter 17).
- The MLEs are consistent and asymptotically normally distributed.

Logit and Probit Models

Example

(From Chapter 14 of Cameron and Trivedi, 2005) Fishing mode choice
The dependent variable is binary with

$$y_i = \begin{cases} 1 & \text{if fishing from a charter boat} \\ 0 & \text{if fishing from a pier} \end{cases}.$$

The single explanatory variable is

$$x_i = \ln(price_{c,i} / price_{p,i}).$$

The prices of charter boat and pier fishing vary across individuals owing to various factors, for example, to differences in access. It is expected that the probability of charter boat fishing decreases as its relative price increases.

Example

(Continued) Regression results are (numbers in parentheses are t-ratios)

Regressor	Logit	Probit	OLS
Constant	2.053 (12.15)	1.194 (13.34)	0.784 (65.38)
ln relp	-1.823 (-12.61)	-1.056 (-13.87)	-0.243 (-28.15)

Coefficient estimates for the price variable are all negative as expected.