Advanced Econometrics

Chapter 14: Discrete Response Models

In Choi

Sogang University

Useful references:
This note heavily depends on Wooldridge (2002).

- Amemiya, T. (1985). Advanced econometrics. Harvard university press.
- Cameron, A.C., and P.K. Trivedi (2005). Microeconometrics: Methods and Applications. Cambridge University Press: New York.
- Maddala, G. S. (1986). Limited-dependent and qualitative variables in econometrics (No. 3). Cambridge university press.
- Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

Introduction

- Suppose that the observed random variable y_i takes either value 0 or 1. Models for this type of data are called binary response models.

Example

$y_i = 0$ if the i-th person is unemployed and $y_i = 1$ if he/she is employed.

- We omit index i in the below.
- Estimating the response probability

$$p(\mathbf{x}) \equiv P(y = 1 \mid \mathbf{x}) = P(y = 1 \mid x_1, ..., x_K)$$

is the main concern of this chapter.

Introduction

- For a continuous variable x_j, the partial effect of x_j on the response probability is
$$\frac{\partial P(y=1 \mid \mathbf{x})}{\partial x_j}.$$
For a binary variable x_K,interest lies in
$$P(y=1 \mid x_1, ..., 1) - P(y=1 \mid x_1, ..., 0).$$

Linear probability model

- The linear probability model (LPM) for y is specified as

$$P(y = 1 \mid \mathbf{x}) = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K.$$

  Here β_1 signifies the change in the probability of success $(y = 1)$ given a one-unit increase in x_1.

- Since y is a binary random variable,

$$\begin{aligned} E(y \quad &\mid \quad \mathbf{x}) = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K; \\ Var(y \quad &\mid \quad \mathbf{x}) = \mathbf{x}'\beta(1 - \mathbf{x}'\beta). \end{aligned}$$

  This is a linear model with heteroskedasticity. Thus, the OLS estimator is consistent and asymptotically normal. White's Robust t-test should be used.

Linear probability model

- If $0 < \hat{y}_i < 1$ for all i, the estimated standard deviation is $\hat{\sigma}_i = \sqrt{\hat{y}_i(1 - \hat{y}_i)}$. The feasible GLS estimator of β is obtained by regressing

$$y_i / \hat{\sigma}_i \text{ on } 1/\hat{\sigma}_i, x_{i1}/\hat{\sigma}_i, ..., x_{iK}/\hat{\sigma}_i.$$

- In practice, $\hat{y}_i$ may be greater than 1 or less than zero, which is awkward because y_i is either 0 or 1.

Index models: Probit and logit

- Consider the model

$$P(y = 1 \mid \mathbf{x}) = G(\mathbf{x}'\beta).$$

- Assume $0 < G(z) < 1$ for all $z \in R$.
- The conditional probability depends on $\mathbf{x}$ only through the index $\mathbf{x}'\beta$.
- In most applications, G is a cdf.

Index models: Probit and logit

- Index models where G is a cdf can be derived from an underlying latent variable model

$$y^* = \mathbf{x}'\beta + e, \quad y = 1(y^* > 0),$$

  where e is a continuously distributed variable independent of x and the distribution of e is symmetric about zero. Here $E(y^* \mid \mathbf{x}) = \beta_0 + \beta_1 x_1 + ... + \beta_K x_K$.
- Then,

$$\begin{aligned} P(y &= 1 \mid \mathbf{x}) \\ &= P(y^* > 0 \mid \mathbf{x}) \\ &= P(e > -\mathbf{x}'\beta \mid \mathbf{x}) \\ &= 1 - G(-\mathbf{x}'\beta) \\ &= G(\mathbf{x}'\beta) \text{ (symmetry of } G\text{)}. \end{aligned}$$

Index models: Probit and logit

- The probit model assumes

$$G(z) = \Phi(z) = \int \phi(v)dv,$$

  where $\phi(z)$ is the standard normal density.
- The logit model assumes a standard logit distribution for e. That is,

$$G(z) = \Lambda(z) = \frac{\exp(z)}{1 + \exp(z)}.$$

Index models: Probit and logit

- Interpretation of β_j

$$\frac{\partial p(\mathbf{x})}{\partial x_j} = g(\mathbf{x}'\beta)\beta_j, \ g(z) = \frac{dG(z)}{dz}.$$

  This shows that the partial effect of x_j on $p(\mathbf{x})$ depends on x through $g(\mathbf{x}'\beta)$. Since $g(z) > 0$ for all z, the sign of the effect is given by the sign of β_j .

- If x_K is a binary variable, the partial effect from changing x_k from zero to one is

$$G(\beta_0 + \beta_1 x_1 + ... + \beta_K) - G(\beta_0 + \beta_1 x_1 + ... + \beta_{K-1} x_{K-1})$$

Maximum likelihood estimation of binary response index models

- Assume we have N iid observations. The density of y_i given x_i can be written as

$$f(y \mid x_i;\beta) = \left[G(\mathbf{x}_i'\beta)\right]^y \left[1 - G(\mathbf{x}_i'\beta)\right]^{1-y}, \; y = 0, 1.$$

- The log-likelihood function is given as

$$L(\beta) = \sum_{i=1}^{N} \left(\log\left[G(\mathbf{x}_i'\beta)\right] + (1 - y_i)\log\left[1 - G(\mathbf{x}_i'\beta)\right]\right).$$

- The usual theory of MLE applies here. The formula for the variance-covariance matrix is

$$Avar(\hat{\beta}) = \left\{\sum_{i=1}^{N} \frac{\left[g(\mathbf{x}_i'\hat{\beta})\right]^2 x_i' x_i}{\left[G(\mathbf{x}_i'\hat{\beta})\right]\left[1 - G(\mathbf{x}_i'\hat{\beta})\right]}\right\}^{-1}.$$

- This can be used for statistical inference as usual.

Tests for binary response index models

- Consider the model

$$P(y = 1 \mid \mathbf{x}, \mathbf{z}) = G(\mathbf{x}'\beta + \mathbf{z}'\gamma).$$

  The null hypothesis of interest is $H_0 : \gamma = 0$.
- Wald, likelihood ratio and Lagrange multiplier tests can be constructed in a usual manner.

Tests for binary response index models

- Consider the model

$$P(y = 1 \mid \mathbf{x}, \mathbf{z}) = G(\mathbf{x}'\beta + \mathbf{z}'\gamma).$$

  The null hypothesis of interest is $H_0 : \gamma = 0$.
- Wald, likelihood ratio and Lagrange multiplier tests can be constructed in a usual manner.

Goodness-of fit for binary response index models

- For each i, compute the predicted probability that $y_i = 1$, given the explanatory variables, x_i. If $G(\mathbf{x}_i'\hat{\beta}) > 0.5$, predict y_i to be unity; otherwise y_i is predicted to be zero. The percentage of times the predicted y_i matches the actual y_i is the percent correctly predicted.
- McFadden suggests a pseudo R-squared

$$1 - L_{ur} / L_0,$$

where L_{ur} is the log-likelihood function for the estimated model and L_0 is that for the model with only an intercept. The pseudo R-squared is always between zero and one.

Measuring marginal effects of regressors

- If x_j is roughly continuous,

$$\Delta P(y=1 \mid \mathbf{x}) \approx \left[g(\mathbf{x}'\hat{\beta})\hat{\beta}_j\right] \Delta x_j$$

  for small changes in x_j. Here any x_i can be used for $g(\mathbf{x}'\hat{\beta})$. Often, in practice, $g(\bar{\mathbf{x}}'\hat{\beta})$ is used for $g(\mathbf{x}'\hat{\beta})$.
- If x_K is a discrete variable, we can estimate the change in the predicted probability in going from c_{K+1} to c_K as

$$G(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \ldots + \hat{\beta}_K c_{K+1}) - G(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \ldots + \hat{\beta}_K c_K).$$

  Note that we are using sample averages for other regressors.

- Consider the model

$$
\begin{aligned}
y_1^* &= z_1'\delta_1 + \alpha_1 y_2 + u_1 \text{ (structural eqn)} \\
y_2 &= z_1'\delta_{21} + z_2'\delta_{22} + v_2 = z'\delta_2 + v_2 \quad \text{(reduced-form eqn)} \\
y_1 &= 1[y_1^* > 0],
\end{aligned}
$$

  where (u_1, v_2) has a zero mean, bivariate normal distribution and is independent of z.

- If the covariance of u_1 and v_2 is non-zero, y_2 is endogenous.

- Rivers and Vuong (1988, JOE) introduce a two-step approach. Under joint normality of (u_1, v_2) with $Var(u_1) = 1$, we can write

$$u_1 = \theta_1 v_2 + e_1,$$

where $\theta_1 = \eta_1/\tau_2^2$, $\eta_1 = Cov(v_2, u_1)$, $\tau_2^2 = Var(v_2)$, and e_1 is independent of z and v_2. We also have $E(e_1) = 0$ and $Var(e_1) = Var(u_1) - \eta_1^2/\tau_2^2 = 1 - \rho_1^2$, where $\rho_1 = Corr(v_2, u_1)$.

- Now write

$$\begin{aligned} y_1^* &= z_1'\delta_1 + \alpha_1 y_2 + \theta_1 v_2 + e_1 \\ e_1 &\mid z, y_2, v_2 \sim N(0, 1 - \rho_1^2). \end{aligned}$$

Trivially,

$$P(y = 1 \mid z, y_2, v_2) = \Phi\left(\left(z_1'\delta_1 + \alpha_1 y_2 + \theta_1 v_2\right) / \sqrt{1 - \rho_1^2}\right).$$

- If v_2 were observed, the standard Probit procedure yields consistent estimates of $\delta_{1\rho} = \delta_1 / \sqrt{1-\rho_1^2}$, $\alpha_{1\rho} = \alpha_1 / \sqrt{1-\rho_1^2}$ and $\theta_{1\rho} = \theta_1 / \sqrt{1-\rho_1^2}$.
- Since v_2 are not observed,
  (i) Run the OLS regression y_2 on z and save the residuals $\hat{v}_2$.
  (ii) Run the probit y_1 on z_1, y_2, $\hat{v}_2$ to get consistent estimators of the scaled coefficients $\delta_{1\rho}$, $\alpha_{1\rho}$, and $\theta_{1\rho}$.

- The joint distribution of (y_1, y_2) conditional on z is

$$f(y_1, y_2) = f(y_1 \mid y_2, z) f(y_2 \mid z).$$

Since $y_2 \mid z \sim N(z\delta_2, \tau_2^2)$ and

$$P(y_1 = 1 \mid y_2, z) = \Phi\left(\frac{z_1'\delta_1 + \alpha_1 y_2 + (\rho_1/\tau_2)(y_2 - z'\delta_2)}{\sqrt{1-\rho_1^2}}\right), \quad (1)$$

we have

$$f(y_1, y_2) = [\Phi(w)]^{y_i}[1-\Phi(w)]^{1-y_i}\frac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\frac{(y_2 - z\delta_2)^2}{2\tau^2}\right),$$

where w denotes the term in inside $\Phi(\cdot)$ in equation (1).

Specification issues
MLE

- The log-likelihood function is written as

$$\sum_{i=1}^{N}\left[y_{i1}\log\Phi(w_i) + (1-y_i)\log\left[1-\Phi(w_i)\right] - \frac{1}{2}\log(\tau^2) - \frac{1}{2}\frac{(y_{2i} - z_i\delta_2}{2\tau^2}\right.$$

  Maximizing this function with respect to the parameters $(\delta_1, \alpha_1, \rho_1, \delta_2, \tau_2)$, we obtain the MLEs.
- The MLEs are consistent and efficient.
- Exogeneity of y_2 can be tested by testing $H_0 : \rho_1 = 0$ using t-test.
- See Wooldridge (2002) for the MLE in the case of binary endogenous explanatory variables.

Extensions

- Multinomial logit (MNL) model
  - The response variable y takes integer values 0,1,...,J.
  - Examples: occupational choice, transportation mode for commuting, etc.
  - Let x be a $K \times 1$ vector with first element being unity. The MNL has response probability

$$
\begin{aligned}
P(y &= j \mid x) = \exp(x'\beta_j)/[1 + \sum_{h=1}^{J} \exp(x'\beta_h)],\ j = 1, ..., J; \\
P(y &= 0 \mid x) = 1/[1 + \sum_{h=1}^{J} \exp(x'\beta_h)].
\end{aligned}
$$

  - MLE is used to estimate the MNL model.

Extensions

- Ordered logit and ordered probit models
  - The response variable y takes integer values $0, 1, \ldots, J$.
  - But the integer values carry some meaning. Examples: credit rating, quintiles of an income distribution
  - MLE is used to estimate the models.