

# Limited Dependent Variable Models

## Semester 2/2014

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
$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$$
  - ▶ How can we interpret  $\beta$  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  $x_j$ 
$$P(y = 1|x) = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k \quad (1)$$
- ▶ Call equation (1) the linear probability model

## Linear probability model

- ▶ The response probability,  $P(y=1|x)$ , is linear in the parameter  $\beta_j$
- ▶  $\beta_j$  measures the change in the probability of success when  $x_j$  changes, holding other factors fixed:  
$$\Delta P(y = 1|x) = \beta_j \Delta x_j$$
- ▶ If the estimated equations is  $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_k x_k$
- ▶  $\hat{y}$  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

# Binary response models

- ▶ The response probability:  $P(y = 1|\mathbf{x}) = P(y = 1|x_1, x_2, \dots, x_k)$
- ▶ In the LPM, we assume that the response probability is linear in a set of parameters.
- ▶ For general setting, define the response probability as a function taking on values between 0 and 1:  $0 < G(z) < 1$   
 $P(y = 1|\mathbf{x}) = G(\beta_0 + \beta_1x_1 + \dots + \beta_kx_k) = G(\beta_0 + \mathbf{x}\beta)$  (1)
- ▶ The two most used function  $G$  that ensure the probabilities between 0 and 1:
  - ▶ the logistic function  $\rightarrow$  logit model
  - ▶ the standard normal cumulative distribution function  $\rightarrow$  probit model

# Binary response models

## Specifying Logit and Probit models

- ▶ Logit model:  $G$  is the logistic function

$$G(z) = \frac{\exp(z)}{1 + \exp(z)} = \Lambda(z) \quad (2)$$

- ▶ Probit model:  $G$  is the standard normal cumulative distribution function (cdf)

$$G(z) = \Phi(z) \equiv \int_{-\infty}^z \phi(v) dv, \quad (3)$$

where  $\phi(z)$  is the standard normal density:

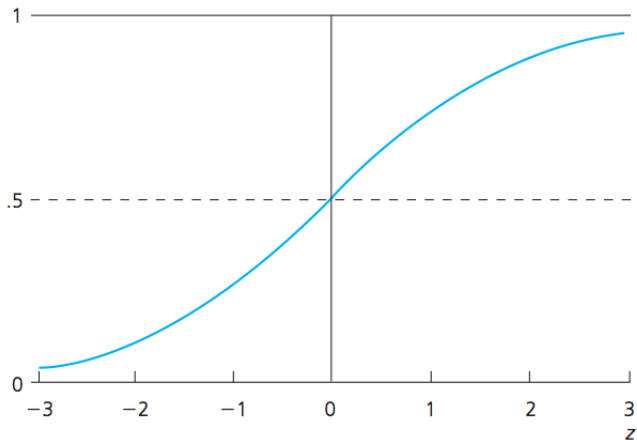
$$\phi(z) = (1/\sqrt{2\pi}) \exp(-z^2/2) \quad (4)$$

- ▶  $G(z)$  in both (2) and (3) are increasing functions with  $G(z) \rightarrow 0$  as  $z \rightarrow -\infty$ , and  $G(z) \rightarrow 1$  as  $z \rightarrow \infty$ .

# Logistic function

Graph of the logistic function  $G(z) = \exp(z)/[1 + \exp(z)]$ .

$$G(z) = \exp(z)/[1 + \exp(z)]$$



## Latent variable model

- ▶ Logit and probit models can be derived from an underlying latent variable model.
- ▶ Let  $y^*$  be an unobserved, or latent, variable determined by  $y^* = \beta_0 + \mathbf{x}\beta + e$ ,  $y = 1[y^* > 0]$  (5)
- ▶  $1[.]$  is an indicator function, which is = 1 if the event in the brackets is true, and = 0 otherwise.
- ▶ Assume  $e$  is independent of  $\mathbf{x}$  and  $e$  either has the standard logistic distribution or the standard normal distribution.
  - ▶  $e$  is symmetrically distributed about zero
  - ▶  $1 - G(-z) = G(z)$  for all real numbers  $z$
- ▶ Probit is more popular since economists tend to favor the normality assumption for  $e$

## Latent variable model

- ▶ From (5), we can derive the response probability for  $y$ :  
$$P(y = 1|\mathbf{x}) = P(y^* > 0|\mathbf{x}) = P[e > -(\beta_0 + \mathbf{x}\beta)|\mathbf{x}]$$
$$= 1 - G[-(\beta_0 + \mathbf{x}\beta)] = G(\beta_0 + \mathbf{x}\beta) \quad (6)$$
- ▶ The direction of the effect of  $x_j$  on  $E(y^* | \mathbf{x}) = \beta_0 + \mathbf{x}\beta$  and on  $E(y|\mathbf{x}) = P(y = 1|\mathbf{x}) = G(\beta_0 + \mathbf{x}\beta)$  is always the same.
- ▶ For the magnitude, we care about the effect of  $x_j$  on the probability of success  $P(y = 1|\mathbf{x})$ .

# Maximum likelihood estimation of logit and probit models

- ▶ Suppose we have a random sample of size  $n$ .
- ▶ To obtain MLE, conditional on the explanatory variables, we need the density of  $y$  given  $\mathbf{x}_j$ :  
$$f(y|\mathbf{x}_j; \beta) = [G(\mathbf{x}_j\beta)]^y [1 - G(\mathbf{x}_j\beta)]^{1-y}, y = 0, 1 \quad (7)$$
- ▶ Then, the log-likelihood function for observation  $i$  is:  
$$\ell_i(\beta) = y_i \log[G(\mathbf{x}_i\beta)] + (1 - y_i) \log[1 - G(\mathbf{x}_i\beta)] \quad (8)$$
- ▶ The log-likelihood for a sample size of  $n$ :  $\ell(\beta) = \sum_{i=1}^n \ell_i(\beta)$   
(9)
- ▶ Maximum likelihood estimator (MLE):  $\hat{\beta}$  maximizes the log-likelihood in (9)
- ▶ Under general conditions, the MLE is consistent, asymptotically normal, and asymptotically efficient.

## Goodness of fit

- ▶ For binary response model, we calculate pseudo R-squared =  $1 - \ell_{ur}/\ell_0$
- ▶  $\ell_{ur}$  = log-likelihood function for the estimated model,  $\ell_0$  = log-likelihood function with only intercept model.
- ▶ The log-likelihood function is always a negative number, and  $|\ell_{ur}| \leq |\ell_0|$ .
- ▶ If the covariates have no explanatory power, then  $\ell_{ur}/\ell_0 = 1$  and pseudo  $R^2 = 0$
- ▶ If  $\ell_{ur} = 0$ , the pseudo  $R^2 = 1$

# Interpretation of probits and logits

- ▶ We care about the effect of  $x$  on  $P(y=1|x)$ , that is  $\partial p/\partial x$ .
- ▶ Probit and logit are non-linear, partial effect on  $p(x)$  is from  $\frac{\partial p(x)}{\partial x_j} = g(\beta_0 + \mathbf{x}\beta)\beta_j$ , where  $g(z) \equiv \frac{dG}{dz}(z)$   
where  $G$  is the cdf,  $g$  is a probability density function
- ▶  $G(\cdot)$  is a strictly increasing cdf, so  $g(z) > 0$  for all  $z$ 
  - ▶ The partial effect always has the same sign as  $\beta_j$
- ▶ So, we can only directly compare sign and significance (based on a standard t test) of coefficients

# Hypotheses testing

- ▶ For each coefficient, we can construct (asymptotic) t tests as usual:  $t = \hat{\beta}_j / se(\hat{\beta}_j)$
- ▶ For multiple hypotheses, or tests of multiple exclusion restrictions, we use the likelihood ratio (LR) test
- ▶ LR test is based on the difference in the log-likelihood functions for the unrestricted and restricted models, whether the log-likelihood becomes significantly smaller (no larger) when dropping variables.
- ▶ LR statistic:  $LR = 2(\ell_{ur} - \ell_r)$
- ▶ Under  $H_0$ , if we are testing  $q$  exclusion restrictions,  $LR \sim \chi_q^2$

# Marginal effects

- ▶ For nonlinear models, the marginal effect (ME) varies with the point of evaluation
- ▶ 3 common choices of evaluation:
  1. AME: average marginal effect (average of ME at each  $x = x_i$  )
  2. MEM: marginal effect at mean (ME at  $x = \bar{x}$  )
  3. MER: marginal effect at a representative value (ME at  $x = x^*$  )
    - ▶ AME: margins,  $dydx(*)$
    - ▶ MEM: margins,  $dydx(*)$  at mean
    - ▶ MER: margins,  $dydx(*)$  at (female =1 age=40)

# Marginal effects

- ▶ If the conditional mean is of the single-index form:  
 $E(y|x) = m(x\beta)$ , then  $ME_j = m'(x\beta) \cdot \beta_j$ 
  - ▶ If  $m(x\beta)$  is monotonically increasing,  $m'(x\beta) > 0$ . The sign of  $\hat{\beta}_j$  gives the sign of the ME.
  - ▶ For any function  $m(\cdot)$  and at any value of  $x$ , we have  $\frac{ME_j}{ME_k} = \frac{\beta_j}{\beta_k}$