Limited Dependent Variable Models I

Fall 2008

Limited Dependent Variables

- A limited dependent variable, Y, is defined as a dependent variable whose range is substantively restricted.

Limited Dependent Variables

- A limited dependent variable, Y, is defined as a dependent variable whose range is substantively restricted.
- The common cases are:

Limited Dependent Variables

- A limited dependent variable, Y, is defined as a dependent variable whose range is substantively restricted.
- The common cases are:
    - binary: $Y \in \{0, 1\}$

Limited Dependent Variables

- A limited dependent variable, Y, is defined as a dependent variable whose range is substantively restricted.
- The common cases are:
  - binary: $Y \in \{0, 1\}$
  - multinomial: $Y \in \{0, 1, 2, ..., k\}$

Limited Dependent Variables

- A limited dependent variable, Y, is defined as a dependent variable whose range is substantively restricted.
- The common cases are:
  - binary: $Y \in \{0, 1\}$
  - multinomial: $Y \in \{0, 1, 2, ..., k\}$
  - integer: $Y \in \{0, 1, 2, ...\}$

Limited Dependent Variables

- A limited dependent variable, Y, is defined as a dependent variable whose range is substantively restricted.
- The common cases are:
  - binary: $Y \in \{0, 1\}$
  - multinomial: $Y \in \{0, 1, 2, ..., k\}$
  - integer: $Y \in \{0, 1, 2, ...\}$
  - censored: $Y \in \{Y^* : Y^* \geq 0\}$

Binary Response Models

- A dependent variable is of qualitative nature, coded as a dummy variable, $Y_i \in \{0, 1\}$.
- Examples:
  - driving to work versus public transportation
  - being a single versus getting married
  - employed vesus unemployed
- We wil analyze two different models
  - linear probability model
  - non-linear models (logit and probit)

Linear Probability Model

- What if we still want to use a multiple linear regression model?

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad Y_i \in \{0, 1\}$$

Linear Probability Model

- What if we still want to use a multiple linear regression model?

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad Y_i \in \{0, 1\}$$

- Because Y can take on only two values, β_j **cannot** be interpreted as the change in Y given one-unit increase of X_j.

Linear Probability Model

- What if we still want to use a multiple linear regression model?

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad Y_i \in \{0, 1\}$$

- Because Y can take on only two values, β_j **cannot** be interpreted as the change in Y given one-unit increase of X_j.
- Under the zero conditional mean assumption,

$$E(Y_i|X) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} = \Pr(Y_i = 1),$$

which is called the *response probability*.

Linear Probability Model

- What if we still want to use a multiple linear regression model?

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad Y_i \in \{0, 1\}$$

- Because Y can take on only two values, β_j **cannot** be interpreted as the change in Y given one-unit increase of X_j.
- Under the zero conditional mean assumption,

$$E\left(Y_i|X\right) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} = \Pr\left(Y_i = 1\right),$$

which is called the *response probability*.
- The multiple linear regression model with a binary dependent variable is called the **linear probability model (LPM)** because the response probability is linear in β_1.

Linear Probability Model

- What if we still want to use a multiple linear regression model?

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i, \quad Y_i \in \{0, 1\}$$

- Because Y can take on only two values, β_j **cannot** be interpreted as the change in Y given one-unit increase of X_j.
- Under the zero conditional mean assumption,

$$E(Y_i|X) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} = \Pr(Y_i = 1),$$

which is called the *response probability*.

- The multiple linear regression model with a binary dependent variable is called the **linear probability model (LPM)** because the response probability is linear in β_1.
- In the LPM, β_j measures the change in the response probability when X_j increases by one unit:

$$\frac{\partial \Pr(Y_i = 1)}{\partial X_j} = \beta_j.$$

Example: Environmental Concern

- In the households survey data in Kuala Lumpur, one question is
  - "Are you concerned about the environment?" "Yes/No"
- The LPM and the estimation are

$$EnvCon_i = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 LogInc_i + \beta_4 Smell_i + u_i.$$

	Age	Sex	Log Income	Smell	Constant
Coeff.	0.008	0.015	0.112	0.130	-0.683
Std. Err.	0.002	0.048	0.030	0.052	0.260

- The predicted probability of concerning about the environment is

$$\widehat{EnvCon}_i = -0.683 + 0.008 Age_i + 0.015 Sex_i + 0.112 LogInc_i + 0.13 Smell_i$$

Limitations of the LPM

- The predicted probability, $\widehat{\beta}_0 + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i}$, can be outside $[0, 1]$.
- (Heteroskedasticity) Conditional on X_js, the residuals take only two values:

$$\begin{aligned} u_i &= 1 - (\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}) \qquad \text{if} \quad Y_i = 1 \\ u_i &= -(\beta_0 + \beta_1 X_i + \beta_2 X_{2i}) \qquad \text{if} \quad Y_i = 0 \end{aligned}$$

Then the variance of the error term is

$$\begin{aligned} Var(u_i|X) &= E(u_i^2|X) \\ &= [1 - \Pr(Y_i = 1)]^2 \Pr(Y_i = 1) \\ &\quad + \Pr(Y_i = 1)^2 [1 - \Pr(Y_i = 1)] \\ &= \Pr(Y_i = 1)[1 - \Pr(Y_i = 1)] \end{aligned}$$

- (Constant marginal effects) Given the linearity of the model,

$$\frac{\partial \Pr(Y_i = 1)}{\partial X_{ji}} = \beta_j.$$

Logit and Probit Models for Binary Response

- The limitations of the LPM can be overcomed by using more sophisticated response models:

$$\Pr(Y_i = 1) = G(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}),$$

where $G(\cdot)$ is a function taking on values between zero and one: $0 < G(z) < 1$ for any real z.

Logit and Probit Models for Binary Response

- The limitations of the LPM can be overcomed by using more sophisticated response models:

$$\Pr\left(Y_i = 1\right) = G\left(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}\right),$$

  where $G\left(\cdot\right)$ is a function taking on values between zero and one: $0 < G\left(z\right) < 1$ for any real z.
- The two commonest functional forms are:

Logit and Probit Models for Binary Response

- The limitations of the LPM can be overcomed by using more sophisticated response models:

$$\Pr(Y_i = 1) = G(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}),$$

  where $G(\cdot)$ is a function taking on values between zero and one: $0 < G(z) < 1$ for any real z.
- The two commonest functional forms are:
  - (logit model) G is the logistic function

$$\Pr(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})}$$

Logit and Probit Models for Binary Response

- The limitations of the LPM can be overcomed by using more sophisticated response models:

$$\Pr(Y_i = 1) = G(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}),$$

  where $G(\cdot)$ is a function taking on values between zero and one: $0 < G(z) < 1$ for any real z.
- The two commonest functional forms are:
  - (logit model) G is the logistic function

$$\Pr(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})}$$

  - (probit model) G is the standard normal cdf

$$\Pr(Y_i = 1) = \Phi(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})$$

Logit and Probit Models for Binary Response

- The limitations of the LPM can be overcomed by using more sophisticated response models:

$$\Pr(Y_i = 1) = G(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}),$$

where $G(\cdot)$ is a function taking on values between zero and one: $0 < G(z) < 1$ for any real z.

- The two commonest functional forms are:
  - (logit model) G is the logistic function

$$\Pr(Y_i = 1) = \frac{\exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})}{1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})}$$

  - (probit model) G is the standard normal cdf

$$\Pr(Y_i = 1) = \Phi(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})$$

- As both models are non-linear, β_j is **not** the marginal effect of X_j on Y, for $j = 1, 2$.

The Structure of the Model

- We define a latent variable, Y_i^*, which is unobservable but determined in the following way:

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i.$$

We observe the variable Y_i which is linked to Y_i^* as

$$Y_i = \begin{cases} 0 & \text{if} \quad Y_i^* \leq 0 \\ 1 & \text{if} \quad Y_i^* > 0 \end{cases}.$$

The Structure of the Model

- We define a latent variable, Y_i^*, which is unobservable but determined in the following way:

$$Y_i^* = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i.$$

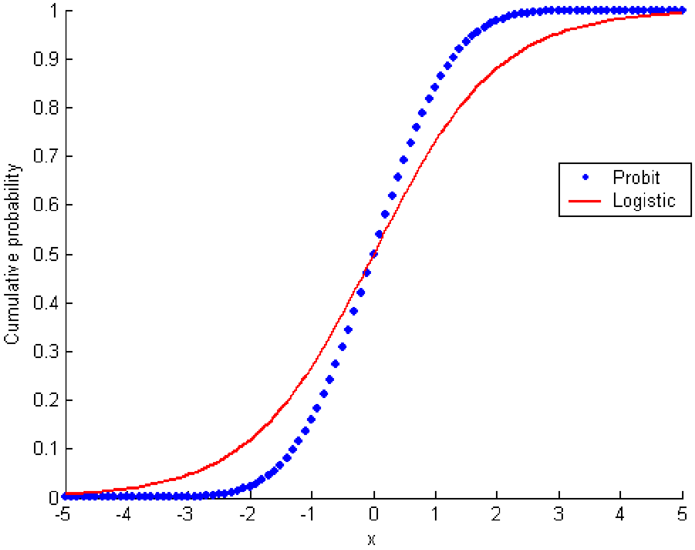
We observe the variable Y_i which is linked to Y_i^* as

$$Y_i = \begin{cases} 0 & \text{if} \quad Y_i^* \leq 0 \\ 1 & \text{if} \quad Y_i^* > 0 \end{cases}.$$

- The probability of observing $Y_i = 1$ is

$$\begin{aligned} \Pr(Y_i = 1) &= \Pr(Y_i^* > 0) \\ &= \Pr(u_i > -\beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i}) \\ &= 1 - G_u(-\beta_0 - \beta_1 X_{1i} - \beta_2 X_{2i}) \\ &= G_u(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}) \end{aligned}$$

Shape of Logit and Probit Models



Marginal Effects

- In most of applications, the primary goal is to explain the effects of X_j on the response probability $\Pr(Y = 1)$.

Marginal Effects

- In most of applications, the primary goal is to explain the effects of X_j on the response probability $\Pr(Y = 1)$.
- Logit Model: for $j = 1, 2$,

$$\frac{\partial \Pr(Y_i = 1)}{\partial X_j} = \frac{\beta_j \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})}{(1 + \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}))^2}$$

$$= \beta_j \Pr(Y_i = 1)(1 - \Pr(Y_i = 1))$$

Thus, one unit increase in X_j leads to an increase of $\beta_j \Pr(Y_i = 1)(1 - \Pr(Y_i = 1))$ in the response probability.

Marginal Effects

- Probit Model:

$$\frac{\partial \Pr\left(Y_i = 1\right)}{\partial X_j} = \beta_j \phi\left(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}\right)$$

Marginal Effects

- Probit Model:

$$\frac{\partial \Pr(Y_i = 1)}{\partial X_j} = \beta_j \phi(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i})$$

- In the both models, the relative effects of any two continuous independent variables, X_1 and X_2, are

$$\frac{\frac{\partial \Pr(Y_i=1)}{\partial X_1}}{\frac{\partial \Pr(Y_i=1)}{\partial X_2}} = \frac{\beta_1}{\beta_2}.$$

Odds-Ratio

- The odds-ratio in a binary response model is defined as $\Pr(Y_i = 1) / [1 - \Pr(Y_i = 1)]$.
- If this ratio is equal to 1, then both outcomes have equal probability. If this ratio is equal to 2 , then the outcome $Y_i = 1$ is twice more likely than the outcome $Y_i = 0$.
- In the logit model, the log odds-ratio is linear in the parameters:

$$\ln\left(\frac{\Pr(Y_i = 1)}{1 - \Pr(Y_i = 1)}\right) = \beta_0 + +\beta_1 X_{1i} + \beta_2 X_{2i}$$

- Thus, in the logit model, β_1 measures the marginal effect of X on the log odds-ratio. That is, a unit increase in X leads to an increase of $100\beta_1\%$ in the odds-ratio.

Maximum Likelihood Estimation I

- Both logit and probit models are non-linear models. We introduce a new estimation method, called the **Maximum Likelihood Estimation**.
- Let $\{(Y_i, X_{1i}, ..., X_{ki})\}_{i=1}^{N}$ denote a random sample from the population distribution of Y conditional on $X_1, ..., X_k$, $f(Y|X_1, ..., X_k; \theta)$.
- The likelihood estimation requires the parametric assumption of functional forms on $f(Y|X_1, ..., X_k; \theta)$ and we need to know the joint distribution.

Maximum Likelihood Estimation II

- Because of the random sampling assumption, the joint distribution of $\{(Y_i, X_{1i}, ..., X_{ki})\}_{i=1}^{N}$ is the product of the distributions:
  - (continuous variable) $\Pi_{i=1}^{n} f\left(Y_i | X_{1i}, ..., X_{ki}; \theta\right)$
  - (discrete variable) $\Pi_{i=1}^{n} \Pr\left(Y_i | X_{1i}, ..., X_{ki}; \theta\right)$
- Then, the *likelihood function* is defined as

$$L\left(\theta; \{(Y_i, X_{1i}, ..., X_{ki})\}_{i=1}^{N}\right) = \Pi_{i=1}^{n} f\left(Y_i | X_{1i}, ..., X_{ki}; \theta\right)$$

- The maximum likelihood (ML) estimator of θ is the value of θ that maximizes the likelihood function.

Maximum Likelihood Estimation III

- The ML principle says that, out of all the possible values for θ, the value that makes the likelihood of the observed data largest should be chosen.
- Usually, it is more convenient to work with the *log-likelihood* function:

$$\log L\left(\theta; \{(Y_i, X_{1i}, ..., X_{ki})\}_{i=1}^{N}\right) = \sum_{i=1}^{n} \log f\left(Y_i | X_{1i}, ..., X_{ki}; \theta\right).$$

- Because of the non-linear nature of the maximization problem, we usually cannot obtain an explict formula for ML estimators. Thus, one requires a numerical optimization.
- Under very general conditions, *the MLE is consistent, asymptotically efficient, and asymptotically normal.*

Estimation of Logit and Probit Models

- Suppose that we observed N_1 number of $Y = 1$ and N_0 number of $Y = 0$, where $N_0 + N_1 = N$.
- The log-likelihood function from this data is

$$\log L(\beta_0, \beta_1) = \sum_{i=1}^{N_1} \log \Pr(Y_i = 1) + \sum_{i=1}^{N_0} \log \Pr(Y_i = 0),$$

  where $\Pr(Y_i = 1)$ is either a normal cdf or logisitic cdf.
- The MLE of β_0 and β_1 are found by maximizing the log-likelihood with respect to β_0 and β_1:

$$\frac{\partial \log L(\beta_0, \beta_1)}{\partial \beta_0} = 0 \text{ and } \frac{\partial \log L(\beta_0, \beta_1)}{\partial \beta_1} = 0.$$

Example: Environmental Concern

- The model is

$$EnvCon_i = \beta_0 + \beta_1 Age_i + \beta_2 Sex_i + \beta_3 LogInc_i + \beta_4 Smell_i + u_i.$$

	LPM		Logit		Probit	
Variable	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err
Age	0.008	0.002	0.032	0.009	0.020	0.005
Sex	0.015	0.048	0.065	0.204	0.040	0.126
Log Income	0.112	0.030	0.480	0.132	0.299	0.081
Smell	0.130	0.052	0.556	0.224	0.349	0.138
Constant	-0.683	0.260	-5.073	1.160	-3.157	0.708
			Some Marginal Effects			
Age	0.008		0.008		0.008	
Log Income	0.112		0.111		0.119	

Hypothesis Testing I

- We want to do hypothesis testing in logit and probit models. There are several common ways to construct test statistics. We will use the *likelihood ratio* (LR) test.
- The LR test is based on the difference in the log-likelihood functions for the unrestricted and restricted models, just as the F test compared the goodness of fit in MLR models.

Hypothesis Testing II

- Let $\mathcal{L}_{ur}$ ($\mathcal{L}_r$) denote the maximized log-likelihood value for the unrestricted (restricted) model. Then the **likelihood ratio statistic** is

$$LR = 2\left(\mathcal{L}_{ur} - \mathcal{L}_r\right) \sim^a \mathcal{X}_q^2,$$

  where q is the number of restrictions in a hypothesis.
- Note that because $\mathcal{L}_{ur} \geq \mathcal{L}_r$, LR is nonnegative.
- Interpretation: The larger the fall of the log-likelihood is after imposing restrictions, the more likely we want to reject the null hypothesis.

Example: Environmental Concern

- We want to test a hypothesis saying there is no difference in environmental concern according to age and sex.

$$H_0 : \beta_1 = 0 = \beta_2 \text{ vs. } H_1 : \text{not } H_0$$

- The likelihood ratio statistics in logit and probit models are

	Logit		Probit	
	Unrestr.	Restr.	Unrestr.	Restr.
Log Likelihood	-278.84	-286.85	-278.78	-286.84
LR statistics	16.03		16.11	

- The critical value of the chi-squrare distribution with 2 degrees of freedom is 5.99. Hence, we can not accept the null.