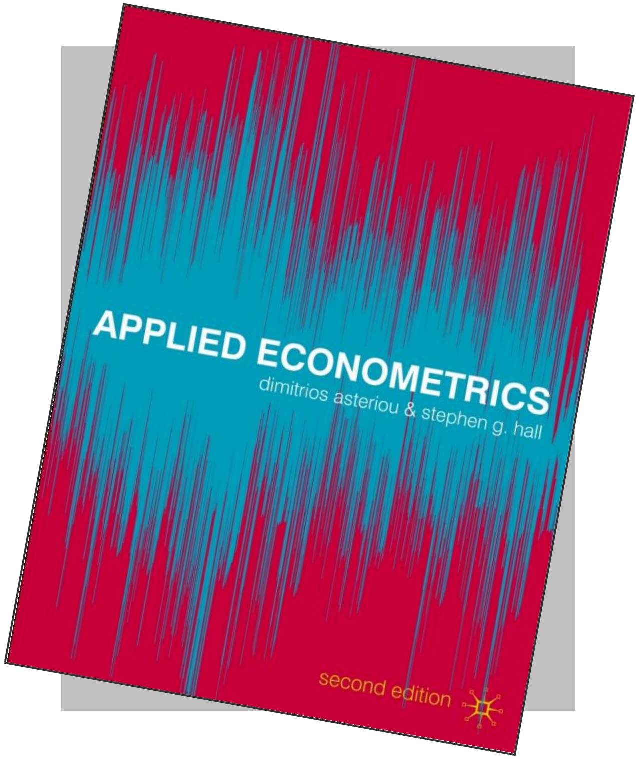
Applied Econometrics
Second edition

Dimitrios Asteriou and
Stephen G. Hall

Chapter 12:
Limited Dependent Variable
Regression Models



palgrave
macmillan

Limited Dependent Variable Regression Models

1. Introduction
2. Linear probability model
3. Problems with the linear probability model
4. Logit model
5. Probit model
6. Tobit model

Learning Objectives

1. Understand problems caused by estimating a model with a dummy dependent variable using simple linear model
2. Be familiar with the logit and probit models for dummy dependent variables
3. Estimate logit and probit models and interpret the results obtained through econometric software
4. Be familiar with and learn how to estimate the multinomial and ordered logit and probit models
5. Understand the meaning of censored data and learn the use of Tobit model
6. Estimate Tobit model for censored data

Introduction

There are frequently cases where a dependent variable is qualitative and therefore a dummy is used on left-hand side of the regression model.

For example, examine why some people go to university and others not, or why some people decide to enter the labour force and others not.

Both these variables are dichotomous (they take 0 or 1 values) dummy variables of the types discussed before.

However, here we want to use this variable as dependent variable.

Linear Probability Model

Examine the simplest possible model, which has a dichotomous dummy variable as the dependent variable.

For simplicity, assume that the dummy dependent variable is explained by only one regressor.

For example, we are interested in examining the labour force participation decision of adult females.

The question is: Why do some women enter the labour force and others not?

For simplicity, assume that the decision to work or not is affected by only one explanatory variable (X_{2i}) – the level of family income.

Linear Probability Model (2)

The model is:

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

As Y_i is a dummy variable, we can rewrite the model as:

$$D_i = \beta_1 + \beta_2 X_{2i} + u_i$$

where X_{2i} is the level of family income (a continuous variable);

D_i is a dichotomous dummy defined as

$$D_i = \begin{cases} 1 & \text{if the } i\text{th individual is working} \\ 0 & \text{if the } i\text{th individual is not working} \end{cases}$$

Linear Probability Model (3)

As D_i is of a qualitative nature, interpretation is different.

Let us define P_i as the probability of $D_i = 1 (P_i = Pr(D_i = 1))$;

Therefore $1-Pi$ is the probability of $D_i = 0$

$(1 - P_i = Pr(D_i = 0))$.

To put this mathematically:

$$\begin{aligned} E(D_i) &= 1 \Pr(D_i = 1) + 0 \Pr(D_i = 0) \\ &= 1P_i + 0(1 - P_i) \\ &= P_i \end{aligned}$$

Linear Probability Model (4)

Therefore, this model simply suggests that the expected value of D_i is equal to the probability that the *ith* individual is working.

For this reason, this model is called the linear probability model.

The values obtained, β_1 and β_2, enable us to estimate the probabilities that a woman with a given level of family income will enter the labour force.

Linear Probability Model (5)

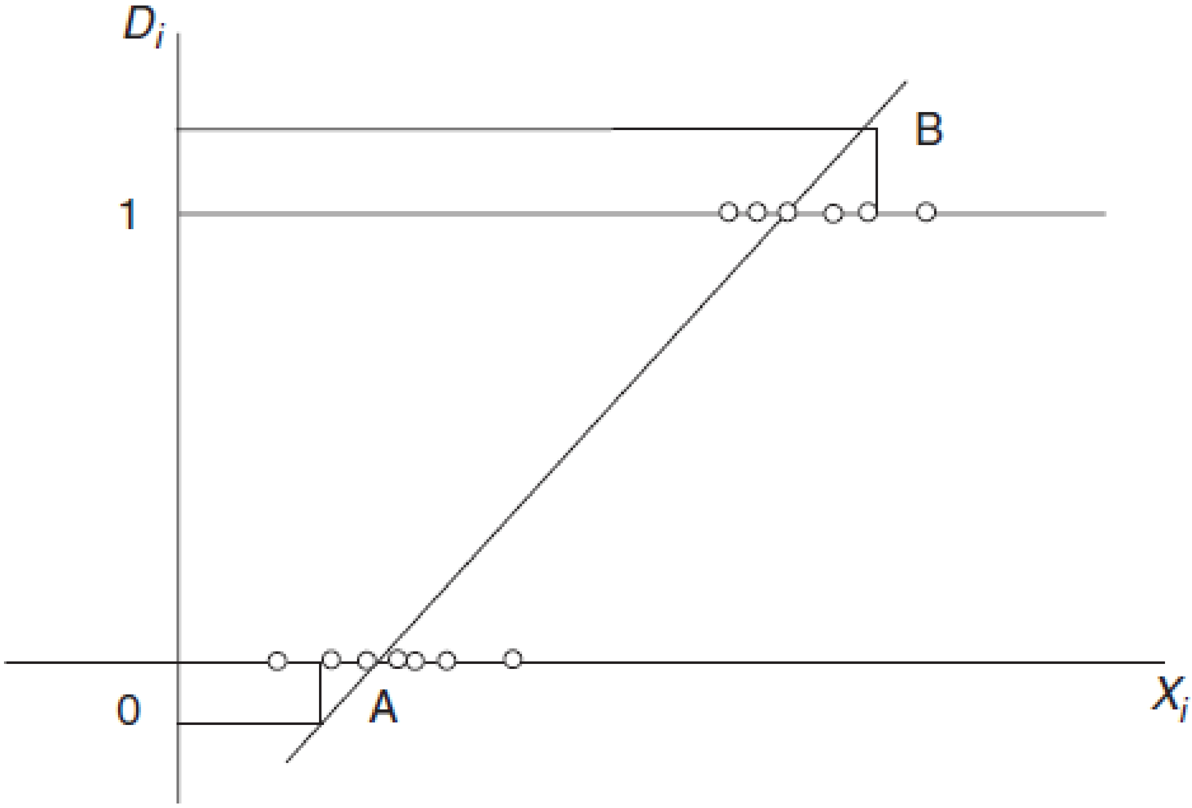


Figure 12.1 The linear probability model

Linear Probability Model (6)

Problems with linear probability model

$\hat{D}_i$ is not bounded by the (0,1) range

Non-normality and heteroskedasticity of the disturbances

Coefficient of determination as measure of overall fit

Logit Model

In the linear probability model, we saw that the dependent variable D_i on the left-hand side, which reflects the probability P_i, can take any real value and is not limited to being in the correct range of probabilities – the (0,1) range.

Logit Model (2)

Simple way to resolve this problem involves two steps. First, transform the dependent variable, D_i, as follows, introducing concept of odds:

$$odds_i = \frac{P_i}{1 - P_i}$$

The second step involves taking the natural logarithm of the odds ratio, calculating the logit, L_i, as:

$$L_i = ln\left(\frac{P_i}{1 - P_i}\right)$$

Logit Model (3)

Using this in a linear regression we obtain the logit model as:

$$L_i = \beta_1 + \beta_2 X_{2i} + u_i$$

This model (which is linear to both the explanatory variable and the parameters) can be extended to more than one explanatory variable, so as to obtain:

$$L_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

Logit Model (4)

Notice that the logit model resolves the 0,1 boundary condition problem because:

(a) As the probability P_i approaches 0 the odds approach zero and the logit *(ln(0))* approaches $-\infty$

(b) As the probability P_i approaches 1 the odds approach $+\infty$ and the logit *(ln(1))* approaches $+\infty$

Logit Model (5)

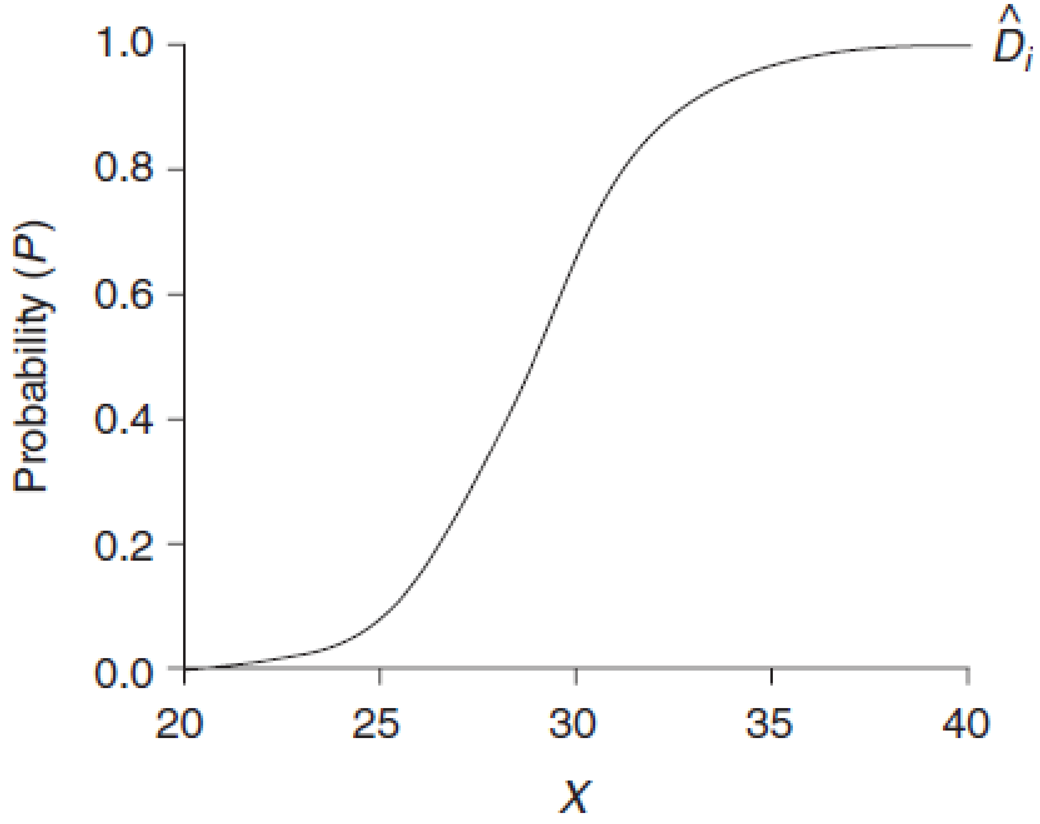


Figure 12.2 The logit function

Logit Model (6)

Interpretation of estimates in logit models

(a) Calculate the change in average D_i

(b) Take the partial derivative

$$\frac{\partial \hat{D}_i}{\partial X_{ji}} = \hat{\beta}_j \hat{D}_i (1 - \hat{D}_i)$$

(c) Multiply the obtained β_j coefficients by 0.25

Logit Model (7)

Goodness of fit in Logit Model

$$count\ R^2 = \frac{number\ of\ correct\ predictions}{number\ of\ observations}$$

$$R_k^2 = \frac{number\ of\ correct\ predictions\ of\ D_i = 1}{number\ of\ observations\ of\ D_i = 1} + \frac{number\ of\ correct\ predictions\ of\ D_i = 0}{number\ of\ observations\ of\ D_i = 0}$$

$$pseudo\text{-}R^2 = 1 - \frac{l_u}{l_R}$$

Probit Model

Using the cumulative normal distribution to model P_i we have:

$$P_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Z_i} e^{-\frac{s^2}{2}} ds$$

where P_i is the probability that the dependent dummy variable $D_i = 1$, $Z_i = \beta_1 + \beta_2 X_{2i}$

(this can be easily extended to *k-variables case) and s is standardized normal* variable.

Probit Model (2)

Generally, logit and probit analyses provide similar results and similar marginal effects, especially for large samples.

However, since the shapes of the tails of the logit andthe two models produce different results in terms of 0 and 1 values in the dependent dummy variable if the sample is unbalanced. probit distributions are different (see figure),

Probit Model (3)

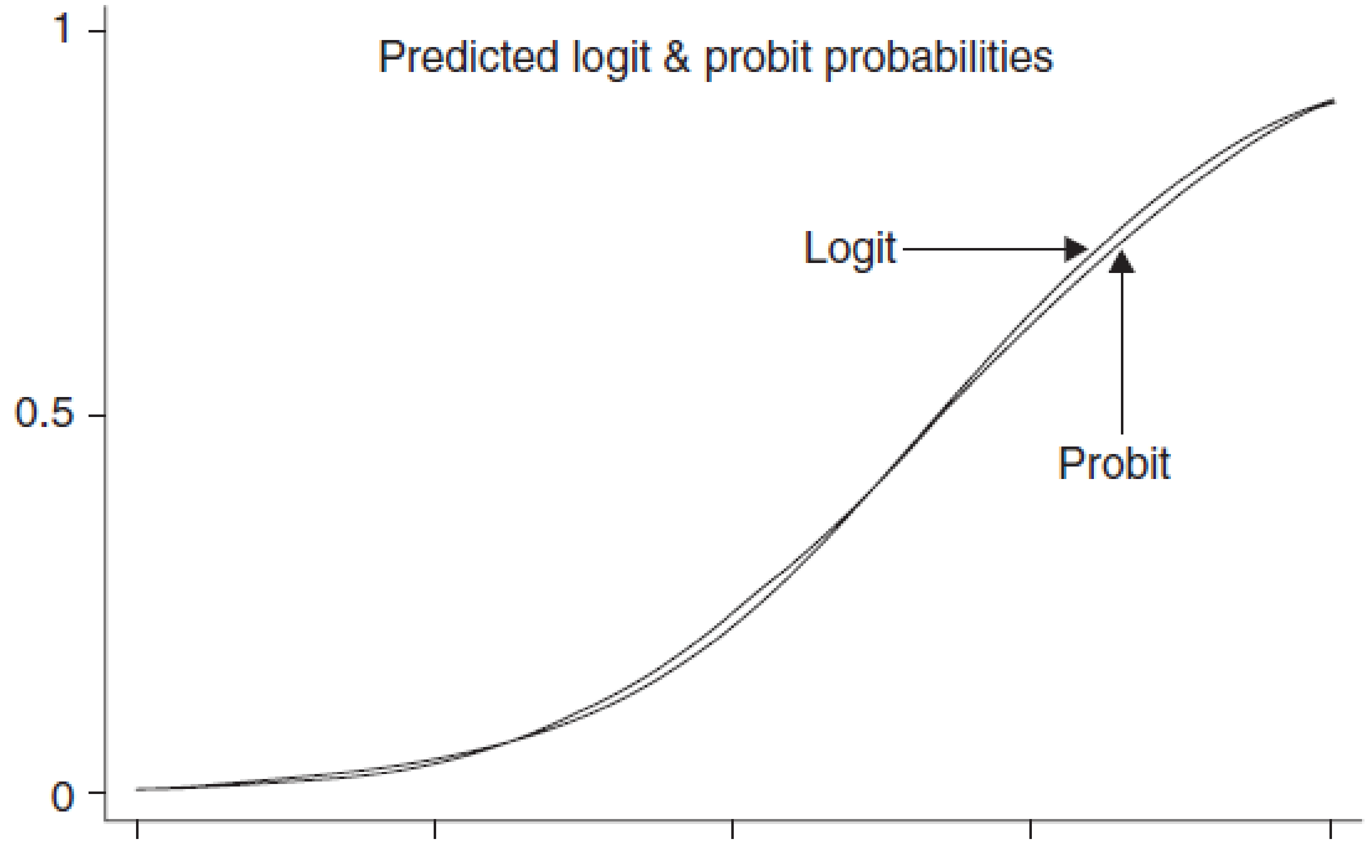


Figure 12.4 Differences between logit and probit probabilities

Applied Econometrics

- Linear probability model

$$p = \mathrm{pr}[y = 1|\mathrm{x}] = \mathrm{x}'\beta \qquad \mathrm{F}(\mathrm{x}'\beta) = \mathrm{x}'\beta$$

- Logit

$$\mathrm{F}(\mathbf{x}'\beta) = \Lambda(\mathbf{x}'\beta) = \frac{e^{\mathbf{x}'\beta}}{1 + e^{\mathbf{x}'\beta}} = \frac{\exp\left(\mathbf{x}'\beta\right)}{1 + \exp\left(\mathbf{x}'\beta\right)}$$

- Probit

$$\mathrm{F}(\mathbf{x}'\beta) = \Phi(\mathbf{x}'\beta) = \int_{-\infty}^{\mathbf{x}'\beta} \phi(z)dz$$

Applied Econometrics

- *Marginal effects at the mean*

$$\partial p/\partial \mathrm{x}_j = \mathrm{F}'(\bar{\mathbf{x}}'\beta)\beta_j$$

- *Average marginal effects*

$$\partial p/\partial \mathrm{x}_j = \frac{\sum \mathrm{F}'(\mathbf{x}'\beta)}{\mathrm{n}}\,\beta_j$$

Multinomial Logit and Probit Models

are multi-equation modelsMultinomial logit and probit models . A dummy dependent variable with *k* categories will create *k* − *1* equations (and cases to examine).

This is easy to see if we consider that for the dichotomous dummy $D = (1, 0)$ we have only one logit/probit equation to capture the probability that the one or the other will be chosen. Therefore, if we have a trichotomous (three different choices) variable we need two equations, and for a *k* categories variable, *k* − *1* equations.

Multinomial Logit and Probit Models (2)

Example

$$D_S = \begin{cases} 1 & \text{if the takeover is by shares} \\ 0 & \text{if otherwise} \end{cases}$$

$$D_C = \begin{cases} 1 & \text{if the takeover is by cash} \\ 0 & \text{if otherwise} \end{cases}$$

$$D_M = \begin{cases} 1 & \text{if the takeover is by a mixture} \\ 0 & \text{if otherwise} \end{cases}$$

$$D_S = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i$$

$$D_C = a_1 + a_2 X_{2i} + a_3 X_{3i} + \cdots + a_k X_{ki} + v_i$$

Ordered Logit and Probit Models

Where multiple response categories of a dummy variable follow a rank or order, as in strongly agree, agree and so on, ordered logit and probit models are used.

These models assume that the observed *Di is determined by* D^*_i using the rule:

$$D_i = \begin{cases} 1 & \text{if } D_i^* \leq \gamma_1 \\ 2 & \text{if } \gamma_1 \leq D_i^* \leq \gamma_2 \\ 3 & \text{if } \gamma_2 \leq D_i^* \leq \gamma_3 \\ & \vdots \\ M & \text{if } \gamma_M \leq D_i^* \end{cases}$$

Tobit Model

Tobit model is an extension of the probit model that allows us to estimate models that use censored variables.

Censored variables are variables that contain regular values for some of the cases in the sample and do not have any values at all in some other cases.

Tobit Model (2)

Take the simple dummy variable of home ownership, which takes the values of:

$$D_i = \begin{cases} 1 & \text{if the } i\text{th individual is home owner} \\ 0 & \text{if the } i\text{th individual is not home owner} \end{cases}$$

However, if we transform this variable to examine the amount of money spent in order to buy a house, we have continuous values for those who own a house and a set of zeros for those individuals who do not own a house. Such variables are called **censored**

Tobit Model (3)

The problem relies on the fact that a simple OLS estimation of models of this kind will ‘ignore’ zero values of censored dependent variable and provide results that are biased and inconsistent.

Tobit model resolves the problem by providing appropriate parameter estimates. The mathematics behind the Tobit model is rather complicated and beyond the scope of this book.