Limited dependent variable models

Mauro Costantini
Brunel University

Table of Contents

Introduction and motivation

- Individuals or firms make choices "either-or" in many situations
- Economics: Why some individuals vote for a increasing public spending and others do not; why some female college students decided to study physics and others do not
- Finance: Why firms decide to list their shares on the NASQAD rather than the NYSE; Why some firms pay dividend while others do not; Why some firms choose to engage in stock splits while others do not

Introduction and motivation

- For all the examples previously shown the dependent variable would be a dummy variable $0 - 1$ since there are only two possible outcomes
- This situation would be regarded to as *limited dependent variables* (Binary choice model)

Binary choice model: an economic example

- How can individual's choice between driving (private transportation) and taking the bus (public transportation) be explained in terms of two possible alternative.
- Let us represent an individual's choice by the following indicator variable:

$$y = \begin{cases} 1 & \text{individual drives to work} \\ 0 & \text{individual takes bus to work} \end{cases}, \tag{1}$$

Binary choice model: an economic example

- If one collect a random sample of workers going to work, then the outcome y will be unknown until the sample is drawn. Therefore, y is a random variable.
- If the probability that an individual drives to work is p, then $P[y = 1] = p$. It follows that the probability that a person uses public transportation is $P[y = 0] = 1 - p$.

Binary choice model: an economic example

- The probability function for such a binary random variable is:

$$f(y) = p^y(1-p)^{1-p},\ y = 0,1 \tag{2}$$

  where p it the probability that y takes value one.

- What factors might affect the probability that an individual chooses one or another solution. One factors may be how long it takes to get work one way or the other:

  $x = (\textit{commuting time by bus} - \textit{communting time by car})$

The linear probability model: an example in Finance

- The linear probability model is by far the simplest way of dealing with binary dependent variables.
- It is based on an assumption that the probability of an event occurring (P_i) is linearly related to a set of explanatory variables $x_{2i}, x_{3i}, ..., x_{ki}$:

$$P_i = p(y_i = 1) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + \mu_i, \; i = 1, .., N. \quad (3)$$

The linear probability model: an example in Finance

- The actual probabilities cannot be observed, so one would estimate a model (with OLS) where y_i (the series of zeros and ones) would be the dependent variable.
- The fitted values from this regression are the estimated probabilities for $y_i = 1$ for each observation i.

The linear probability model: an example in Finance

- Suppose, for example, that we wanted to model the probability that a firm i will pay a dividend ($y_i = 1$) as a function of its market capitalisation (x_{2i}, measured in millions of US dollar)
- The slope estimates for the linear probability model can be interpreted as the change in the probability that the dependent variable will equal 1 for a one-unit change in a given explanatory variable, holding the effect of all other explanatory variables fixed.

The linear probability model: an example in Finance

- Suppose we fit the following equation regression :

$$\hat{P}_i = -0.3 + 0.012x_{2i} \tag{4}$$

where $\hat{P}_i$ denotes the fitted or estimated probability for firm i

- The previous model says that for every \$1m increase in size, the probability that the firm will pay a dividend increases by 0.012 (or 1.2%). A firm whose stock is valued at \$50m will have a $-0.3 + 0.012 \times 50 = 0.3$ (or 30%) probability of making a dividend payment.

The linear probability model: an example in Finance

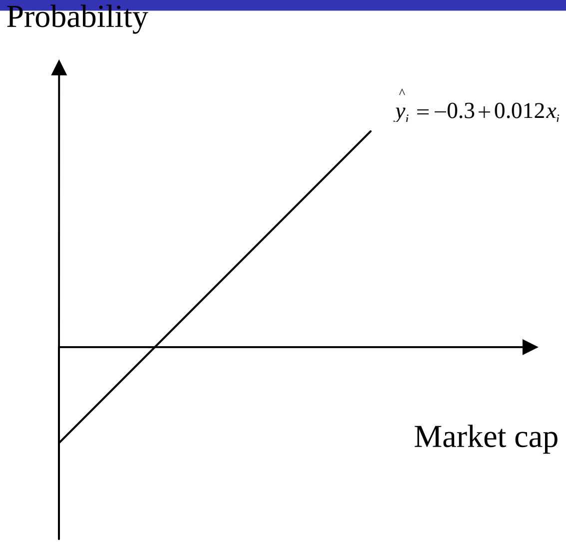

Figure 1. Dividend and market capitalization

The linear probability model: an example in Finance

- if one looks at diagram, then it appears quite clear that some problems emerge. For any firm whose value is less than \$25m, the model predicts a negative probability of dividend payment, while for any firm whose value is greater than \$88m, the probability is greater than one.
- Such predictions cannot obviously be allowed to stand, since the probabilities should lie within the range (0,1). Once can truncate the probabilities at 0 or 1, so that a probability of -0.3, say, would be set to zero, and a probability of, say, 1.2 would be set to 1. However, there are at least two reasons why this is still not adequate.

The linear probability model: an example in Finance

- There will be too many observations for which the estimated probability is zero or one.
- It is not plausible that the probability of paying a dividend is zero or 1. Are we sure that very small firms never pay a dividend and large ones always make a payout?

The linear probability model: cope with heteroscedasticity

- Since the dependent variable takes only one or two values, for given values of independent variables, the error term will also takes one or two values. Consider the following equation:

$$y_i = \beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + \mu_i, \tag{5}$$

- If $y_i = 1$,then by definition, we have

$$\mu_i = 1 - \beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} \tag{6}$$

- if $y_i = 0$,then by definition, we have

$$\mu_i = -\beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} \tag{7}$$

The linear probability model: cope with heteroscedasticity

- Since μ_i changes systematically with the independent variables, then the error term will be also heteroscedastic. therefore, Estimate equation (5) by OLS. Then estimate the variance of the error term: $\sigma_i^2 = \widehat{\mu_i}$.
- Using this variance, transform the data as follows: $y_i^* = y_i/\hat{\sigma}_i$, $x_{2i}^* = x_{2i}/\hat{\sigma}_i$, $x_{3i}^* = x_{3i}/\hat{\sigma}_i$ and so on. Then estimate the following model using the OLS method:

$$y_i^* = \beta_1 \hat{\sigma}_i^{-1} + \beta_2 x_{2i}^* + ... + \beta_k^* x_{ki} + \mu_i^* \tag{8}$$

The probit model

- Consider the following model:

$$P_i = p(y_i = 1) = \beta_1 + \beta_2 x_{2i} + \mu_i, \tag{9}$$

- The linear probability model implicitly assumes that increases in x have a constant effect on the probability, That is, as x_{2i} increases the probability continues to increase at a constant rate:

$$\frac{dp}{dx} = \beta_2 \tag{10}$$

However, since $0 \leq p \leq 1$, a constant rate of increase is impossible.

The probit model: s-shaped curve

- To overcome the problem related to the linear probability model, we consider the probit model.
- To keep the choice probability p within the interval [0,1], a S-shaped relationship between x and p can be used.

The probit model: standard normal distribution

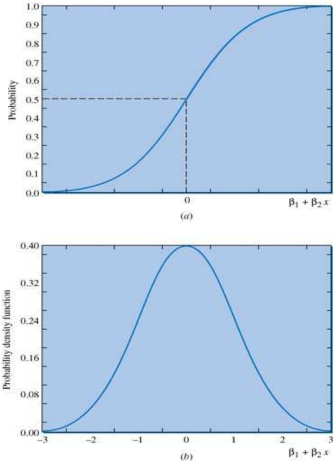

Figure 2 a) Standard normal cumulative distribution function; b) Standard normal probability density function

The probit model

- As x increases, the probability curve rises rapidly at first, and then begins to increase at a decreasing rate. The slope of this curve gives the change in probability given a unit change in x. The slope is not constant as in the linear probability model
- A functional relationship to represent such a curve is the probit function. The probit function is associated with the standard normal distribution

The probit model: probit function

- If Z is a standard normal variable, then its probability density function is:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-1/2} z^2 \qquad (11)$$

- The probit function is:

$$\phi(z) = PZ \leq z = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-1/2} u^2 du \qquad (12)$$

- This integral expression is the probability that a standard normal random variable falls to the left of point z. The function $\phi(z)$ is the cumulative distribution function (cdf) that

The probit model

- Why use the cumulative normal distribution?
- The S-shape gives us what we want:

$$0 \leq Pr(y_i = 1|x_i) \leq, \textit{ for all } x_i$$

$$Pr(y_i = 1|x_i) \textit{ to be increasing in } x_i \textit{ (for } \beta_1 > 0) \tag{13}$$

- Easy to use: The probabilities of the Normal distribution are tabulated.

The probit Model: interpretation

- Consider a probit regression model with one independent variables

$$Pr(y_i = 1) = P[Z \leq \beta_1 + \beta_2 x_{2i}] = \phi(\beta_1 + \beta_2 x_{2i}) \qquad (14)$$

where $\phi(z)$ is the probit function and z the standard normal variable.

- The marginal effect of a one-unit change in x on the probability that $y_i = 1$ is

$$\frac{dp}{dx} = \frac{d\Phi(t)}{dt} \times \frac{dt}{dx} = \phi(\beta_1 + \beta_2 x_{2i})\beta_2 \qquad (15)$$

where $t = \beta_1 + \beta_2 x_{2i}$ and $\Phi(\beta_1 + \beta_2 x_{2i})$ is the standard normal probability density function evaluated at $\beta_1 + \beta_2 x_{2i}$. We estimate this effect by replacing the unknown parameters by their estimates $\hat{\beta}_1$ and $\hat{\beta}_2$.

The probit model: interpretation

- In Figure 2 we show the probit function $\phi(\mathbf{z})$ and the standard normal probability density function $\phi(z)$ just below it. The expression in (15) shows the effect of an increase in x on p. The effect depends on the slope of the probit function, which is given by $\phi(\beta_1 + \beta_2 x_{2i})$ and the magnitude of the parameter β_2.
- Equation (15) has the following implications: Since $\phi(\beta_1 + \beta_2 x_{2i})$ is a probability density function, its value is always positive. Consequently the sign of dp/dx is determined by the sign of β_2. In the transportation problem, we expect β_2 to be positive so that $dp = dx > 0$; as x increases, we expect p to increase.

The probit model: interpretation

- As x changes, the value of the function $\phi(\beta_1 + \beta_2 x_{2i})$ changes. The standard normal probability density function reaches its maximum when $z = 0$ or when $\beta_1 + \beta_2 x_{2i} = 0$. This implies $p = \phi(0) = 0.5$; an individual is equally likely to choose car or bus transportation: the effect of a change in x has its greatest effect, since the individual is "on the borderline" between car and bus transportation. The slope of the probit function $p = \phi(z)$ is at its maximum when $z = 0$, the borderline case.
- On the other hand, if $\beta_1 + \beta_2 x_{2i}$ is large ($\simeq 3$), then the probability that the individual chooses to drive is very large and close to one. This implies that a change in x_{2i} will have relatively little effect, since $\phi(\beta_1 + \beta_2 x_{2i})$ will be nearly zero. The same is true if $\beta_1 + \beta_2 x_{2i}$ is a large negative value ($\simeq -3$)

The probit model: an example with mortgage data.

- Regress mortgage denial (deny) on the payment to-income ratio (P/I)

$$\widehat{P(deny=1)} = \phi(\underset{(0.16)}{-2.19} + \underset{(0.47)}{2.97} P/I) \tag{16}$$

- The estimated coefficient of -2.19 and 2.97 are difficult to interpret because they affect the probability of denial via the $z - value$. What we can say is that the P/I is positively related to probability of denial (the coefficient is positive) and that this relationship is statistically significant (t=2.97/0.47=6.32).

The probit model: An example with mortgage data.

- What is the change in the predicted probability that an application will be denied when P/I increase from 0.3 to 0.4? Compute the probability for $P/I = 0.3$ and for $P/I = 0.4$, and then compute the difference.
- The probability of denial when $P/I = 0.3$ is $\phi(-2.19 + 2.97 \times 0.3) = \phi(-1.30) = 0.097$. The probability of denial when $P/I = 0.4$ is $\phi(-2.19 + 2.97 \times 0.4) = \phi(-1.00) = 0.159$.
- The estimated change in probability of denial is 0.159-0.097=0.062: An increase in payment-to-income ratio from 0.3 to 0.4 is associated with an increase in the probability of denial of 6.2 percentage points, from 9.7% to 15.9%

The logit model

- A frequently alternative for binary choices is the the logit model. These models differ only in the particular S-shaped curve used to constrain probabilities to the [0,1] interval. If L is a logistic random variable, then its probability density function is

$$\lambda(l) = \frac{e^{-l}}{(1+e^{-l})^2}, -\infty < l < \infty \tag{17}$$

- The cumulative distribution function for a logistic random variable is

$$\lambda(l) = P[L \leq l_i] = \frac{1}{(1+e^{-l_i})} \tag{18}$$

The logit model

- Consider a model with only one independent variable. The probability that $y = 1$ can be written:

$$p = \frac{1}{1 + e^{-(\beta_1 + \beta_2 x_{2i})}} \tag{19}$$

- The probability that $y = 0$ can be written:

$$1 - p = \frac{1}{1 + e^{(\beta_1 + \beta_2 x_{2i})}} \tag{20}$$

The logit model

- With the logic model, 0 and 1 are asymptotes to the function and thus the probabilities will never actually fall to exactly zero or rise to one, although they may come infinitesimally close.
- In equation (18), as I_i tends to infinity, e^{-I_i} tends to zero and $1/(1+e^{-I_i})$ tends to 1; as I_i tends to minus infinity, e^{-I_i} tends to infinity and $1/(1+e^{-I_i})$ tends to 0.

The logit model: testing the pecking order hypothesis

- The theory of firm financing suggests that corporations should use the cheapest methods of financing their activities first (i.e. the sources of funds that require payment of the lowest rates of return to investors) and switch to more expensive methods only when the cheaper sources have been exhausted. This is known as the "pecking order hypothesis".
- Differences in the relative cost of the various sources of funds are argued to arise largely from information asymmetries: the firm's senior managers will know the true riskiness of the business, whereas potential outside investors will not. Hence, all else equal, firms will prefer internal finance and then, if further (external) funding is necessary, the firm's riskiness will determine the type of funding sought. The more risky the firm is perceived to be the less accurate will be the pricing of its securities.

The logit model: testing the pecking order hypothesis

- Helwege and Liang (1996) examine the pecking order hypothesis in the context of a set of US firms that had been newly listed on the stock market in 1983, with their additional funding decisions being tracked over the 1984-1992 period. Such newly listed firms are argued to experience higher rates of growth, and are more likely to require additional external funding than firms which have been stock market listed for many years.
- The list of initial public offerings (IPOs) came from the Securities Data Corporation and the Securities and Exchange Commission with data obtained from Compustat. A core objective of the paper is to determine the factors that affect the probability of raising external financing.

The logit model: testing the pecking order hypothesis

- The dependent variable is a binary one: 1 (firm raises funds externally), 0 (firm does not raise any external funds).
- The independent variables are a set that try to grasp the relative degree of information asymmetry and degree of riskiness of the firm. If the pecking order hypothesis is supported by the data, then firms should be more likely to raise external funding the less internal cash they hold. Therefore, the variable "deficit" measures (capital expenditures + acquisitions + dividends - earnings).

The logit model: testing the pecking order hypothesis

- "Positive deficit" is a variable identical to deficit but with any negative deficits (i.e. surpluses) set to zero. "Surplus" is equal to the negative of deficit for firms where deficit is negative." Positive deficit x operating income" is an interactio term where the two variables are multiplied together to capture cases where firms have strong investment opportunities but limited access to internal funds
- "Assets is used as a measure of firm size. "Industry asset growth" is the average rate of growth of assets in that firm's industry over the 1983-1992 period."Firm's growth of sales" is the growth rate of sales averaged over the previous 5 years. "Previous financing" is a dummy variable equal to 1 for firms that obtained external financing in the previous year.

The logit model: testing the pecking order hypothesis.

Table: Logit estimation of the probability of external funding

Variable	(1)	(2)	(3)
Intercept	−0.29 [−3.42]	−0.72 [−7.05]	−0.15 [−1.58]
Deficit	0.04 [0.34]	0.02 [0.18]	
Positive deficit		−0.24 [−1.19]	
Surplus		−2.06 [−3.23]	
Positive deficit x operating income		−0.03 [−0.59]	
Assets	0.0004 [1.99]	0.0003 [1.36]	0.0004 [1.99]
Industry	−0.002 [−1.70]	−0.002 [−1.35]	−0.002 [−1.69]
Previous financing		0.79 [8.48]	

Notes: t-ratio in parenthesis.

The logit model: testing the pecking order hypothesis.

- The key variable, "deficit," has a parameter that is not statistically significant and hence the probability of obtaining external financing does not depend on the size of a firm's cash deficit.
- The parameter on the "surplus" variable has the correct negative sign, indicating that the larger a firm's surplus, the less likely it is to seek external financing, which provides some limited support for the pecking order hypothesis. Larger firms (with larger total assets) are more likely to use the capital markets, as are firms that have already obtained external financing during the previous year.

Logit or probit models

- For the majority of the applications, the logit and probit models will give very similar characterizations of the data because the densities are very similar.
- That is the relationships between the explanatory variables and the probability that $y_i = 1$ will also be very similar. Both approaches are much preferred to the linear probability model.
- The only instance where the models may give non-negligibility different results occurs when the split of the y_i between 0 and 1 is very unbalanced – for example, when $y_i = 1$ occurs only 10% of the time.