Chapter 12

Limited Dependent Variable Models

Some Examples of when Limited Dependent Variables may be used

- There are numerous examples of instances where this may arise, for example where we want to model:
- Why firms choose to list their shares on the NASDAQ rather than the NYSE
- Why some stocks pay dividends while others do not
- What factors affect whether countries default on their sovereign debt
- Why some firms choose to issue new stock to finance an expansion while others issue bonds
- Why some firms choose to engage in stock splits while others do not.
- It is fairly easy to see in all these cases that the appropriate form for the **dependent variable would be a 0-1 dummy variable** since there are only two possible outcomes. There are, of course, also situations where it would be more useful to allow the dependent variable to take on other values, but these will be considered later.

1. The Linear Probability Model

- We will first examine a simple and obvious, but unfortunately flawed, method for dealing with binary dependent variables, known as the **linear probability model.**
- It is based on an assumption that the probability of an event occurring, p_i, is linearly related to a set of explanatory variables

$$y_i = p_i + u_i$$
$$p_i = P(y_i = 1) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \ldots + \beta_k x_{ki}$$

- The **actual probabilities cannot be observed**, so we would estimate a model where the outcomes, y_i (the series of zeros and ones), would be the dependent variable.
- This is then a **linear regression model** and would be estimated by OLS.
- The set of explanatory variables could include either quantitative variables or dummies or both.
- The fitted values from this regression are the estimated probabilities for $y_i = 1$ for each observation i.

The Linear Probability Model

- 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.
- Suppose, for example, that we wanted to model the **probability that a firm *i* will pay a dividend** $p(y_i = 1)$ **as a function of its market capitalisation** (x_{2i}, measured in millions of US dollars), and we fit the following line:

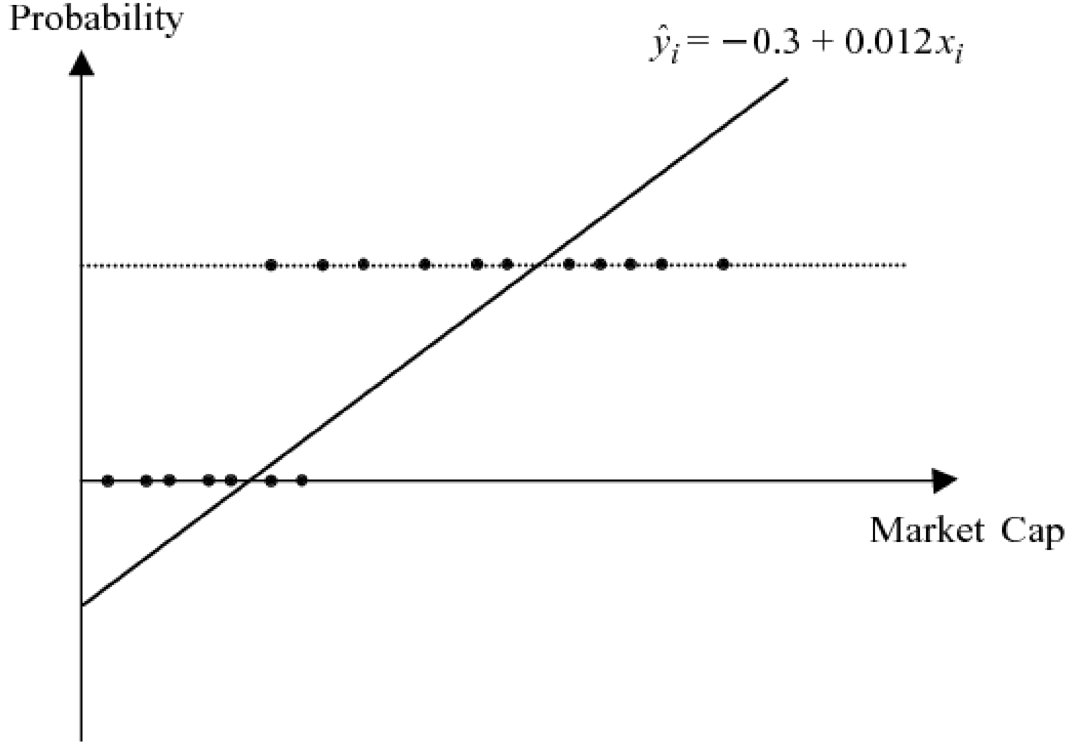
$$\hat{P}_i = -0.3 + 0.012 x_{2i}$$

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

- This model suggests 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 $-0.3+0.012\times50=0.3$ (or 30%) probability of making a dividend payment.

The Fatal Flaw of the Linear Probability Model

- Graphically, the situation we have is



Disadvantages of the Linear Probability Model

- While the linear probability model is simple to estimate and intuitive to interpret, the diagram on the previous slide should immediately signal a problem with this setup.

- For any firm whose value is less than \$25m, the model-predicted **probability** of dividend payment is **negative**, while for any firm worth more than \$88m, the probability is **greater than one**.

- Clearly, such predictions cannot be allowed to stand, since the probabilities should lie within the range (0,1).

- An **obvious solution is to 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.

Disadvantages of the Linear Probability Model 2

- However, there are at least two reasons why this is still not adequate.
- The process of truncation will result in **too many observations for which the estimated probabilities are exactly zero or one**.
- More importantly, it is simply not plausible to suggest that the firm's probability of paying a dividend is either exactly zero or exactly one. Are we really certain that very small firms will definitely never pay a dividend and that large firms will always make a payout?
- Probably not, and so a different kind of model is usually used for binary dependent variables either a **logit or a probit** specification.

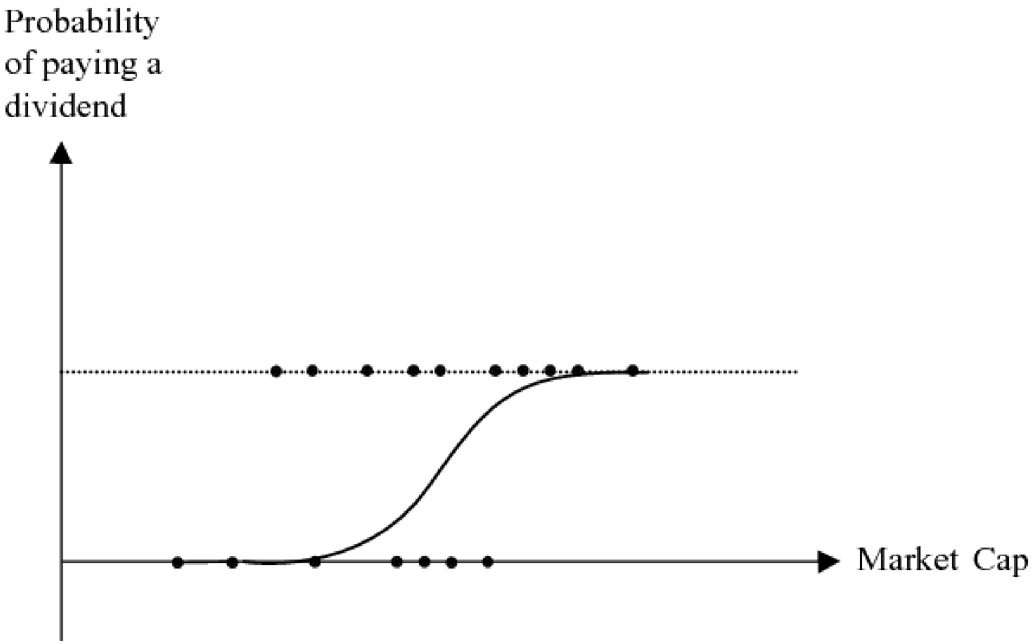
Disadvantages of the Linear Probability Model 3

- The LPM also suffers from a couple of more standard econometric problems that we have examined in previous chapters.

- Since the **disturbance term** changes systematically with the explanatory variables, the former **will also be heteroscedastic**.
  - It is therefore essential that heteroscedasticity-robust standard errors are always used in the context of limited dependent variable models.

Logit and Probit: Better Approaches

- Both the logit and probit model approaches are able to overcome the limitation of the LPM that it can produce estimated probabilities that are negative or greater than one.
- They do this by using a function that effectively transforms the regression model so that the fitted values are bounded within the (0,1) interval.
- Visually, the fitted regression model will appear as an S-shape rather than a straight line, as was the case for the LPM.



2. The Logit Model

- The logit model is so-called because it uses a the **cumulative logistic distribution** to transform the model so that the probabilities follow the S-shape given on the previous slide.

- With the logistic 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.

- The logit model is **not linear** (and cannot be made linear by a transformation) and thus is not estimable using OLS.

- Instead**, maximum likelihood** is usually used to estimate the parameters of the model.

Example: Using a Logit to Test 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 then only switch to more expensive methods 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 since 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.

Data

- 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.
  - They are also more likely to exhibit information asymmetries due to their lack of a track record.

- The **list of initial public offerings (IPOs)** was obtained from the Securities Data Corporation and the Securities and Exchange Commission with data obtained from Compustat.

Aims of the Study and the Model

- A core objective of the paper is to determine the **factors that affect the probability of raising external financing**.
- As such, the **dependent variable will be binary** -- that is, a column of 1's (firm raises funds externally) and 0's (firm does not raise any external funds).
- Thus OLS would not be appropriate and hence a logit model is used.
- The **explanatory variables are a set that aims to capture 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.

Variables used in the Model

- The variable ***deficit*** measures (capital expenditures + acquisitions + dividends - earnings).
- ***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 × operating income*** is an interaction 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 one for firms that obtained external financing in the previous year.

Results from Logit Estimation

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

Source: Helwege and Liang (1996)

Analysis of Results

- 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.
- Or an alternative explanation, as with a similar result in the context of a standard regression model, is that the probability varies widely across firms with the size of the cash deficit so that the standard errors are large relative to the point estimate.
- 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.

3. The Probit Model

- Instead of using the cumulative logistic function to transform the model, the **cumulative normal distribution** is sometimes used instead.
- This gives rise to the probit model.
- As for the logistic approach, this function provides a **transformation to ensure that the fitted probabilities will lie between zero and one**.

Logit or Probit?

- 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 fitted regression plots will be virtually indistinguishable, and the implied 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.
- **Stock and Watson (2006) suggest that the logistic approach was traditionally preferred since the function does not require the evaluation of an integral and thus the model parameters could be estimated faster**.
- However, this argument is no longer relevant given the computational speeds now achievable and **now the choice of one specification rather than the other is now usually arbitrary.**

Parameter Interpretation for Logit and Probit Models

- Standard errors and *t*-ratios will automatically be calculated by the econometric software package used, and hypothesis tests can be conducted in the usual fashion.
- However, interpretation of the coefficients needs slight care.
- It is tempting, but incorrect, to state that a 1-unit increase in x_{2i}, for example, causes a β_2 % **increase in the probability** that the outcome corresponding to y_i =1 will be realized.
- This would have been the **correct interpretation for the linear probability model**.
- However, for logit or probit models, this interpretation would be incorrect because the form of the function is $P_i = \beta_1 + \beta_2 x_{2i} + u_i$, for example, but rather $\mathrm{P_i} = \Phi(z)$ where Φ represents the (non-linear) logistic or cumulative normal function.

Parameter Interpretation for Logit and Probit Models

- To obtain the required relationship between changes in x_{2i} and P_i, we would need to differentiate Φ with respect to x_{2i} and it turns out that this derivative is $\beta_2 F(x_{2i})$.
- So in fact, a 1-unit increase in x_{2i} will cause a $\beta_2 F(x_{2i})$ increase in probability.
- Usually, these impacts of incremental changes in an explanatory variable are evaluated by setting each of them to their mean values.
- These estimates are sometimes known as the **marginal effects**.
- There is also another way of interpreting discrete choice models known as the random utility model.
- The idea is that we can view the value of *y* that is chosen by individual *i* (either 0 or 1) as giving that person a particular level of utility, and the choice that is made will obviously be the one that generates the highest level of utility.
- This interpretation is particularly useful in the situation where the person faces a choice between more than 2 possibilities – see a later slide.

Goodness of Fit for Probit and Logit Models

- While it would be possible to calculate the values of the standard goodness of fit measures such as *RSS*, R^2 these cease to have any real meaning.
- R^2, if calculated in the usual fashion, will be misleading because the fitted values from the model can take on any value but the actual values will only be either 0 and 1.
- Thus if $y_i = 1$ and $\hat{P}_i = 0.8$, the model has effectively made the correct prediction, whereas R^2 and will not give it full credit for this.
- **Two goodness of fit** measures that are commonly reported for limited dependent variable models are
  - The **percentage of y_i values correctly predicted**
  - A measure known as '**pseudo-R^2**' (also known as McFadden's R^2), defined as one minus the ratio of the LLF for the logit or probit model to the LLF for a model with only an intercept.

4. Multinomial Linear Dependent Variables

- There are many instances where investors or financial agents are faced with **more alternatives than a simple binary choice**.

- For example:
  - A company may be considering listing on the **NYSE, the NASDAQ or the AMEX markets**.
  - A firm that is intending to take over another may choose to **pay by cash, with shares, or with a mixture of both.**
  - A retail investor may be choosing between **5 different mutual funds**.
  - A credit ratings agency could assign 1 of 16 (AAA to B3/B-) **different ratings classifications to a firm's debt**.

Multinomial Linear Dependent Variables (Cont'd)

- Notice that the first three of these examples are different from the last one.

- **In the first three cases, there is no natural ordering of the alternatives**: the choice is simply made between them.

- **In the final case, there is an obvious ordering**, because a score of 1, denoting a AAA-rated bond, is better than a score of 2, denoting a AA1/AA+-rated bond, and so on.

- These two situations need to be distinguished and a different approach used in each case. In the first (when there is no natural ordering), a multinomial logit or probit would be used, while in the second (where there is an ordering), an ordered logit or probit would be used.

Discrete Choice Problems

- When the alternatives are unordered, this is sometimes called a discrete choice or multiple choice problem.
- The models used are derived from the **principles of utility maximisation - that is, the agent chooses the alternative that maximises his utility relative to the others.**
- Econometrically, this is captured using a simple generalisation of the binary setup discussed earlier. Thus the multinomial logit and probit are direct extensions of their binary counterparts.
- When there were only 2 choices (0, 1), we required just one equation to capture the probability that one or the other would be chosen.
- If there are now three alternatives, we would need two equations; for four alternatives, we would need three equations. In general, **if there are *m* possible alternative choices, we need *m*-1 equations**.

Example: Modelling the Travel to Work Choice

- The multiple choice example most commonly used is that of the selection of the mode of transport for travel to work.
- Suppose that the journey may be made by **car, bus, or bicycle** (3 alternatives), and suppose that the explanatory variables are the person's **income** (*I*), total **hours worked** (*H*), their **gender** (*G*) and the **distance travelled** (*D*).
- We could set up **2 equations** (e.g., for bus and car) and then travel by bicycle becomes a sort of reference point.
- While the fitted probabilities will always sum to unity by construction, as with the binomial case, there is no guarantee that they will all lie between zero and one.
- In order to make a prediction about which mode of transport a particular individual will use, given that the parameters in, the **largest fitted probability would be set to one and the others set to zero**.

5. Ordered Response Models

- Some limited dependent variables can be assigned numerical values that have a **natural ordering**.
- The most common example in finance is that of **credit ratings**, as discussed previously, but a further application is to modelling a security's bid-ask spread.
- In such cases, it would not be appropriate to use multinomial logit or probit since these techniques cannot take into account any ordering in the dependent variables.
- Using the credit rating example, the model is set up so that a particular bond falls in the AA+ category (using Standard and Poor's terminology) if its unobserved (latent) **creditworthiness** falls within a certain range that is too low to classify it as AAA and too high to classify it as AA.
- The boundary values between each rating are then estimated along with the model parameters.

Are Unsolicited Credit Ratings Biased Downwards?

- The main credit ratings agencies construct ***solicited ratings***, which are those where the issuer of the debt contacts the agency and pays them a fee for producing the rating.
- Many firms globally do not seek a rating (because, for example, the firm believes that the ratings agencies are not well placed to evaluate the riskiness of debt in their country or because they do not plan to issue any debt or because they believe that they would be awarded a low rating).
- But the agency may produce a rating anyway. Such 'unwarranted and unwelcome' ratings are known as ***unsolicited ratings***.
- All of the major ratings agencies produce unsolicited ratings as well as solicited ones, and they argue that there is a market demand for this information even if the issuer would prefer not to be rated.
- Companies in receipt of unsolicited ratings argue that these are biased downwards relative to solicited ratings, and that they cannot be justified without the level of detail of information that can only be provided by the rated company itself.

Data and Methodology

- A study by Poon (2003) seeks to test the conjecture that **unsolicited ratings are biased** after controlling for the rated company's characteristics that pertain to its risk.
- The data employed comprise a pooled sample of all companies that appeared on the annual ‘issuer list’ of S&P during the 1998-2000 years.
- This list contains both solicited and unsolicited ratings covering 295 firms over 15 countries and totaling 595 observations.
- **As expected, the financial characteristics of the firms with unsolicited ratings are significantly weaker than those for firms that requested ratings**.
- The core methodology employs an ordered probit model with explanatory variables comprising firm characteristics and a dummy variable for whether the firm's credit rating was solicited or not:

$$R_i^* = X_i\beta + \epsilon_i \qquad \text{with} \qquad R_i = \begin{cases} 1 & \textit{if } R_i^* \leq \mu_0 \\ 2 & \textit{if } \mu_0 < R_i^* \leq \mu_1 \\ 3 & \textit{if } \mu_1 < R_i^* \leq \mu_2 \\ 4 & \textit{if } \mu_2 < R_i^* \leq \mu_3 \\ 5 & \textit{if } R_i^* > \mu_3 \end{cases}$$

Methodology Continued

where

- R_i are the observed ratings scores that are given numerical values as follows: AA or above = 6, A = 5, BBB = 4, BB = 3, B = 2 and CCC or below = 1
- R_i^* is the unobservable 'true rating' (or 'an unobserved continuous variable representing S&P's assessment of the creditworthiness of issuer *i*')
- X_i is a vector of variables that explain the variation in ratings
- β is a vector of coefficients; μ_i are the threshold parameters to be estimated
- ε_i is a disturbance term that is assumed normally distributed.

- The explanatory variables attempt to capture the creditworthiness using publicly available information.

Definitions of Variables

- Two specifications are estimated: the first includes the variables listed below, while the second additionally incorporates an interaction of the main financial variables with a dummy variable for whether the firm's rating was solicited (SOL) and separately with a dummy for whether the firm is based in Japan.

- The Japanese dummy is used since a disproportionate number of firms in the sample are from this country.

- The financial variables are ICOV - interest coverage (i.e. earnings \ interest); ROA - return on assets; DTC - total debt to capital; and SDTD - short term debt to total debt.

- Three variables SOVAA, SOVA, and SOVBBB are dummy variables that capture the debt issuer's sovereign credit rating (AA; A; BBB or below)

Ordered Probit Results for the Determinants of Credit Ratings

Explanatory Variables	Model 1 Coefficient	Model 1 Test statistic	Model 2 Coefficient	Model 2 Test Statistic
Intercept	2.324	8.960***	1.492	3.155***
SOL	0.359	2.105**	0.391	0.647
JP	-0.548	-2.949***	1.296	2.441**
JP∗SOL	1.614	7.027***	1.487	5.183***
SOVAA	2.135	8.768***	2.470	8.975***
SOVA	0.554	2.552**	0.925	3.968***
SOVBBB	-0.416	-1.480	-0.181	-0.601
ICOV	0.023	3.466***	-0.005	-0.172
ROA	0.104	10.306***	0.194	2.503**
DTC	-1.393	-5.736***	-0.522	-1.130
SDTD	-1.212	-5.228***	0.111	0.171
SOL∗ICOV	–	–	0.005	0.163
SOL∗ROA	–	–	-0.116	-1.476
SOL∗DTC	–	–	0.756	1.136
SOL∗SDTD	–	–	-0.887	-1.290
JP∗ICOV	–	–	0.009	0.275
JP∗ROA	–	–	0.183	2.200**
JP∗DTC	–	–	-1.865	-3.214***
JP∗SDTD	–	–	-2.443	-3.437***
AA or above	>5.095		>5.578	
A	> 3.788 and ≤5.095	25.278***	> 4.147 and ≤5.578	23.294***
BBB	>2.550 and ≤3.788	19.671***	>2.803 and ≤4.147	19.204***
BB	>1.287 and ≤2.550	14.342***	>1.432 and ≤2.803	14.324***
B	>0 and ≤1.287	7.927***	> 0 and ≤1.432	7.910***
CCC or below	≤0		≤0	

Source: Poon (2003)