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

Some Examples of when Limited Dependent Variables may be used (Cont'd)

- 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.

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

$$P_i = p(y_i = 1) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i$$

- 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 Linear Probability Model (Cont'd)

- 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:

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

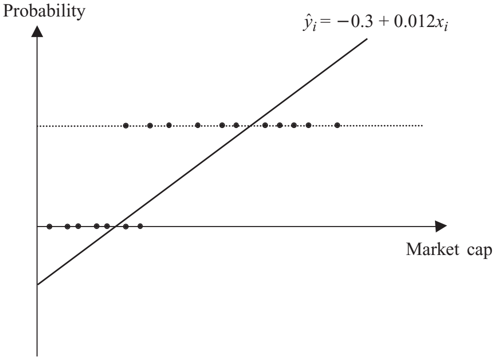
where $\hat{P}_i$ denotes the fitted or estimated probability for firm i.

The Linear Probability Model (Cont'd)

- 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 $-0.3+0.012 50=0.3$ (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

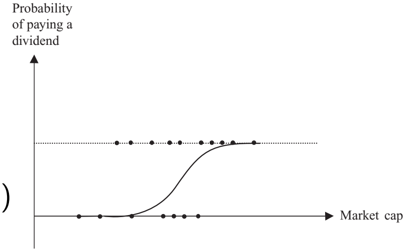
- The LPM also suffers from a couple of more standard econometric problems that we have examined in previous chapters.

- Since the dependent variable only takes one or two values, for given (fixed in repeated samples) values of the explanatory variables, the disturbance term will also only take on one of two values.
  - Hence the error term cannot plausibly be assumed to be normally distributed.

- 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.



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.

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

Variables used in the Model (Cont'd)

- *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 × 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)	

Note: a blank cell implies that the particular variable was not included in that regression; t-ratios in parentheses; only figures for all years in the sample are presented.
Source: Helwege and Liang (1996). Reprinted with the permission of Elsevier.

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.

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 characterisations 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.

Logit or Probit? (Cont'd)

- 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 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 $\beta_2\%$ increase in the probability that the outcome corresponding to $y_i = 1$ will be realised.
- 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 $P_i = F(x_{2i})$ where F 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 F 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*.

Parameter Interpretation for Logit and Probit Models (Cont'd)

- 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.

Parameter Estimation for Probit and Logit Models

- Given that both logit and probit are non-linear models, they cannot be estimated by OLS.
- Maximum likelihood (ML) is invariably used in practice
- The principle is that the parameters are chosen to jointly maximise a log-likelihood function (LLF)
- The form of this LLF will depend upon whether it is the logit or probit model
- While t-test statistics are constructed in the usual way, the standard error formulae used following the ML estimation are valid asymptotically only

Parameter Estimation for Probit and Logit Models (Cont'd)

- Consequently, it is common to use the critical values from a normal distribution rather than a *t*-distribution with the implicit assumption that the sample size is sufficiently large
- For the logit model, assuming that each observation on $y\ i$ is independent, the joint likelihood will be the product of all N marginal likelihoods
- Let $L\left(\theta \left| x_{2i}, x_{3i}, \ldots, x_{ki}; i = 1, N\right.\right)$ denote the likelihood function of the set of parameters $(\beta_1, \beta_2, \ldots, \beta_k)$ given the data.

The Likelihood Function for Probit and Logit Models

- Then the likelihood function will be given by

$$L(\theta) = \prod_{i=1}^{N} \left(\frac{1}{1 + e^{-z_i}}\right)^{y_i} \times \left(\frac{1}{1 + e^{z_i}}\right)^{(1-y_i)}$$

- It is computationally much simpler to maximise an additive function of a set of variables than a multiplicative function
- We thus take the natural logarithm of this equation and so log-likelihood function is maximised

$$LLF = -\sum_{i=1}^{N} [y_i \ln(1 + e^{-z_i}) + (1 - y_i) \ln(1 + e^{z_i})]$$

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.

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.

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.

Are Unsolicited Credit Ratings Biased Downwards? (Cont'd)

- 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.

Data and Methodology (Cont'd)

- 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:

$$R_i^* = X_i\beta + \epsilon_i$$

with

$$R_i = \begin{cases} 1 & \text{if } R_i^* \leq \mu_0 \\ 2 & \text{if } \mu_0 < R_i^* \leq \mu_1 \\ 3 & \text{if } \mu_1 < R_i^* \leq \mu_2 \\ 4 & \text{if } \mu_2 < R_i^* \leq \mu_3 \\ 5 & \text{if } R_i^* > \mu_3 \end{cases}$$

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	

Note: *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.
Source: Poon (2003). Reprinted with the permission of Elsevier.

Analysis of Ordered Probit Results

- The key finding is that the SOL variable is positive and statistically significant in Model 1 (and it is positive but insignificant in Model 2).
- This indicates that even after accounting for the financial characteristics of the firms, unsolicited firms receive ratings on average 0.359 units lower than an otherwise identical firm that had requested a rating.
- The parameter estimate for the interaction term between the solicitation and Japanese dummies (SOL*JP) is positive and significant in both specifications, indicating strong evidence that Japanese firms soliciting ratings receive higher scores.
- On average, firms with stronger financial characteristics (higher interest coverage, higher return on assets, lower debt to total capital, or a lower ratio of short term debt to long term debt) have higher ratings.

The Heckman 2-Step Procedure

- A major flaw that potentially exists within the above analysis is the *self-selection bias* or *sample selection bias* that may have arisen if firms that would have received lower credit ratings (because they have weak financials) elect not to solicit a rating.
- If the probit equation for the determinants of ratings is estimated ignoring this potential problem and it exists, the coefficients will be inconsistent.
- To get around this problem and to control for the sample selection bias, Heckman (1979) proposed a 2-step procedure.

The Heckman 2-Step Procedure (Cont'd)

- In this case would involve first estimating a 0-1 probit model for whether the firm chooses to solicit a rating and second estimating the ordered probit model for the determinants of the rating. The first stage probit model is

$$Y_i^* = Z_i \gamma + \xi_i$$

- where $Y_i = 1$ if the firm has solicited a rating and 0 otherwise, and Y_i^* denotes the latent propensity of issuer i to solicit a rating, Z_i are the variables that explain the choice to be rated or not, and γ are the parameters to be estimated.

The Heckman 2-Step Procedure

- When this equation has been estimated, the rating R_i as defined above in will only be observed if $Y_i = 1$.
- The error terms from the two equations, ϵ_i and ξ_i follow a bivariate standard normal distribution with correlation $\rho_{\epsilon\xi}$.
- The table on the following page shows the results from the two-step estimation procedure, with the estimates from the binary probit model for the decision concerning whether to solicit a rating in panel A and the determinants of ratings for rated firms in panel B.

The Heckman 2-Step Procedure: Results

Explanatory variable	Coefficient	Test statistic
Panel A: Decision to be rated		
Intercept	1.624	3.935***
JP	−0.776	−4.951***
SOVAA	−0.959	−2.706***
SOVA	−0.614	−1.794*
SOVBBB	−1.130	−2.899***
ICOV	−0.005	−0.922
ROA	0.051	6.537***
DTC	0.272	1.019
SDTD	−1.651	−5.320***
Panel B: Rating determinant equation		
Intercept	1.368	2.890***
JP	2.456	3.141***
SOVAA	2.315	6.121***
SOVA	0.875	2.755***
SOVBBB	0.306	0.768
ICOV	0.002	0.118
ROA	0.038	2.408**
DTC	−0.330	−0.512
SDTD	0.105	0.303
JP∗ICOV	0.038	1.129
JP∗ROA	0.188	2.104**
JP∗DTC	−0.808	−0.924
JP∗SDTD	−2.823	−2.430**
Estimated correlation	−0.836	−5.723***
AA or above	>4.275	
A	>2.841 and ≤4.275	8.235***
BBB	>1.748 and ≤2.841	9.164***
BB	>0.704 and ≤1.748	6.788***
B	>0 and ≤0.704	3.316***
CCC or below	≤0	

Note: *, ** and *** denote significance at the 10%, 5% and 1% levels respectively.
Source: Poon (2003). Reprinted with the permission of Elsevier.

The Heckman 2-Step Procedure: Analysis

- A positive parameter value in panel A indicates that higher values of the associated variable increases the probability that a firm will elect to be rated.

- Of the four financial variables, only the return on assets and the short term debt as a proportion of total debt have correctly signed and significant (positive and negative respectively) impacts on the decision to be rated.

- The parameters on the sovereign credit rating dummy variables (SOVAA, SOVA and SOVB) are all significant and negative in sign, indicating that any debt issuer in a country with a high sovereign rating is less likely to solicit its own rating from S&P, other things equal.

The Heckman 2-Step Procedure: Analysis (Cont'd)

- These sovereign rating dummy variables have the opposite sign in the ratings determinant equation (panel B) as expected, so that firms in countries where government debt is highly rated are themselves more likely to receive a higher rating.

- Of the four financial variables, only ROA has a significant (and positive) effect on the rating awarded.

- The dummy for Japanese firms is also positive and significant, and so are three of the four financial variables when interacted with the Japan dummy, indicating that S&P appears to attach different weights to the financial variables when assigning ratings to Japanese firms compared with comparable firms in other countries.

The Heckman 2-Step Procedure: Analysis (Cont'd)

- Finally, the estimated correlation between the error terms in the decision to be rated equation and the ratings determinant equation, $\rho_{\epsilon\xi}$, is significant and negative (-0.836), indicating that the results in table 11.3 above would have been subject to self-selection bias and hence the results of the two-stage model are to be preferred.

Censored and Truncated Variables

- Censored or truncated variables occur when the range of values observable for the dependent variables is limited for some reason.
- Unlike the types of limited dependent variables examined so far, censored or truncated variables may not necessarily be dummies.
- A standard example is that of charitable donations by individuals.
- It is likely that some people would actually prefer to make negative donations (that is, to receive from the charity rather than to donate it), but since this is not possible, there will be many observations at exactly zero.

Censored and Truncated Variables (Cont'd)

- So suppose, for example that we wished to model the relationship between donations to charity and peoples' annual incomes, in pounds.
- Given the observed data, with many observations on the dependent variable stuck at zero, OLS would yield biased and inconsistent parameter estimates.
- An obvious, but flawed, way to get around this would be just to remove all of the zero observations altogether, since we do not know whether they should be truly zero or negative.
- However, as well as being inefficient (since information would be discarded), this would still yield biased and inconsistent estimates.

Censored and Truncated Variables (Cont'd)

- This arises because the error term in such a regression would not have an expected value of zero, and it would also be correlated with the explanatory variable(s).
- For both censored and truncated data, OLS will not be appropriate, and an approach based on maximum likelihood must be used, although the model in each case would be slightly different.
- We can work out the marginal effects given the estimated parameters, but these are now more complex than in the logit or probit cases.

The Differences between Censored and Truncated Variables

- When the terms are used in econometrics, censored and truncated data are different
- Censored data occur when the dependent variable has been ‘censored’ at certain point so that values above (or below) this cannot be observed.
- Even though the dependent variable is censored, the corresponding values of the independent variables are still observable.

The Differences between Censored and Truncated Variables (Cont'd)

- As an example, suppose that a privatisation IPO is heavily oversubscribed, and you were trying to model the demand for the shares using household income, age, education, and region of residence as explanatory variables. The number of shares allocated to each investor may have been capped at, say 250, resulting in a truncated distribution.

- In this example, even though we are likely to have many share allocations at 250 and none above this figure, all of the observations on the independent variables are present and hence the dependent variable is censored, not truncated.

Truncated Variables

- A truncated dependent variable, on the other hand, occurs when the observations for both the dependent and the independent variables are missing when the dependent variable is above (or below) a certain threshold.
- Thus the key difference from censored data is that we cannot observe the x_Is either, and so some observations are completely cut out or 'truncated' from the sample.

Truncated Variables (Cont'd)

- For example, suppose that a bank were interested in determining the factors (such as age, occupation and income) that affected a customer's decision as to whether to undertake a transaction in a branch or on-line. Suppose also that the bank tried to achieve this by encouraging clients to fill in an on-line questionnaire when they log on. There would be no data at all for those who opted to transact in person since they probably would not have even logged on to the bank's web-based system and so would not have the opportunity to complete the questionnaire.

Truncated Variables (Cont'd)

- Thus, dealing with truncated data is really a sample selection problem because the sample of data that can be observed is not representative of the population of interest - the sample is biased, very likely resulting in biased and inconsistent parameter estimates.

- This is a common problem, which will result whenever data for buyers or users only can be observed while data for non-buyers or non-users cannot.

- Of course, it is possible, although unlikely, that the population of interest is focused only on those who use the internet for banking transactions, in which case there would be no problem.

The Tobit Model

- The approach usually used to estimate models with censored dependent variables is known as tobit analysis, named after Tobin (1958).

- To illustrate, suppose that we wanted to model the demand for privatisation IPO shares, as discussed above, as a function of income (x_{2i}), age (x_{3i}), education (x_{4i}), and region of residence (x_{5i}). The model would be

$$
\begin{aligned}
y_i^* &= \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + u_i \\
y_i &= y_i^* \quad \text{for } y_i^* < 250 \\
y_i &= 250 \quad \text{for } y_i^* \geq 250
\end{aligned}
$$

The Tobit Model (Cont'd)

- y_i^* represents the true demand for shares (i.e. the number of shares requested) and this will only be observable for demand less than 250.

- It is important to note in this model that β_2, β_3, etc., represent the impact on the number of shares demanded (of a unit change in x_{2i}, x_{3i}, etc.) and not the impact on the actual number of shares that will be bought (allocated).

Limitations of the Tobit Model

- Before moving on, two important limitations of tobit modelling should be noted.
- First, such models are much more seriously affected by non-normality and heteroscedasticity than are standard regression models, and biased and inconsistent estimation will result.
- Second, the tobit model requires it to be plausible that the dependent variable can have values close to the limit.
- There is no problem with the privatisation IPO example discussed above since the demand could be for 249 shares.

Limitations of the Tobit Model (Cont'd)

- However, it would not be appropriate to use the tobit model in situations where this is not the case, such as the number of shares issued by each firm in a particular month.
- For most companies, this figure will be exactly zero, but for those where it is not, the number will be much higher and thus it would not be feasible to issue, say, 1 or 3 or 15 shares.
- In this case, an alternative approach should be used.

Models for Truncated Dependent Variables

- For truncated data, a more general model is employed that contains two equations - one for whether a particular data point will fall into the observed or constrained categories and another for modelling the resulting variable.
- The second equation is equivalent to the tobit approach.
- This two-equation methodology allows for a different set of factors to affect the sample selection (for example the decision to set up internet access to a bank account) from the equation to be estimated (for example, to model the factors that affect whether a particular transaction will be conducted on-line or in a branch).

Models for Truncated Dependent Variables (Cont'd)

- If it is thought that the two sets of factors will be the same, then a single equation can be used and the tobit approach is sufficient.

- In many cases, however, the researcher may believe that the variables in the sample selection and estimation equations should be different.

- Thus the equations could be

$$\begin{aligned} a_i^* &= \alpha_1 + \alpha_2 z_{2i} + \alpha_3 z_{3i} + \cdots + \alpha_m z_{mi} + \varepsilon_i \\ y_i^* &= \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i \end{aligned}$$

- where $y_i = y_i^*$ for $a_i^* > 0$ and y_i is unobserved for $a_i^* \leq 0$.

Models for Truncated Dependent Variables (Cont'd)

- a_i^* denotes the relative 'advantage' of being in the observed sample relative to the unobserved sample.
- The first equation determines whether the particular data point i will be observed or not, by regressing a proxy for the latent (unobserved) variable, a_i^*, on a set of factors, z_i.
- The second equation is similar to the tobit model.
- Ideally, the two equations will be fitted jointly by maximum likelihood.
- This is usually based on the assumption that the error terms, are multivariate normally distributed and allowing for any possible correlations between them.

Models for Truncated Dependent Variables (Cont'd)

- However, while joint estimation of the equations is more efficient, it is computationally more complex and hence a two-stage procedure popularised by Heckman (1976) is often used.
- The Heckman procedure allows for possible correlations between the error terms while estimating the equations separately in a clever way.