Profitability	1.09**	1.09**	1.21***	2.16	0.91
	(2.18)	(2.14)	(2.81)	(0.75)	(0.29)
Interest margin	1.66***	1.90***	2.71***	−3.42	−2.84
	(2.90)	(3.41)	(4.96)	(1.18)	(0.94)
Observations	1003	1003	770	233	233
No. of banks	247	247	184	82	82
Hausman test statistic	0.66	0.94	0.76	0.58	0.92
R^2	0.28	0.33	0.30	0.46	0.47

Notes: t-ratios in parentheses. Intercept and country dummy parameter estimates are not shown. Empty cells occur when a particular variable is not included in a regression.

Source: de Haas and van Lelyveld (2006). Reprinted with the permission of Elsevier.

The main result is that during times of banking disasters, domestic banks significantly reduce their credit growth rates (i.e. the parameter estimate on the *crisis* variable is negative for domestic banks), while the parameter is close to zero and not significant for foreign banks. There is a significant negative relationship between home country GDP growth, but a positive relationship with host country GDP growth and credit change in the host country. This indicates that, as the authors expected, when foreign banks have fewer viable lending opportunities in their own countries and hence a lower opportunity cost for the loanable funds, they may switch their resources to the host country. Lending rates, both at home and in the host country, have little impact on credit market share growth. Interestingly, the greenfield and takeover variables are not statistically significant (although the parameters are quite large in absolute value), indicating that the method of investment of a foreign bank in the host country is unimportant in determining its credit growth rate or that the importance of the method of investment varies widely across the sample, leading to large standard errors. A weaker parent bank (with higher loss provisions) leads to a statistically significant contraction of credit in the host country as a result of the reduction in the supply of available funds. Overall, both home-related (‘push’) and host-related (‘pull’) factors are found to be important in explaining foreign bank credit growth.

11.8 Panel Unit Root and Cointegration Tests

11.8.1 Background and Motivation

The principle of unit root testing in the panel context is very similar to that employed in single equations framework discussed in Chapter 8. We noted there that unit root tests of the Dickey–Fuller and Phillips–Perron types have low power, especially for modest sample sizes. This provides a key motivation for using a panel – the hope that more powerful versions of the tests can be employed when time series and crosssectional information is combined – as a result of the increase in sample size. Of course, it would be easier to increase the number of observations by simply increasing the length of the sample period, but this data may not be available, or may be of limited use because of structural breaks in the time series.

While the single series and panel approaches to unit root and stationarity testing appear very similar on the surface, in fact a valid construction and application of the test statistics is much more complex for panels than for single series. One complication arises since different asymptotic distributions for the test statistics may result depending on whether N is fixed and T tends to infinity, or vice versa, or both T and N increase simultaneously in a fixed ratio.

Two important issues to consider are first, that the design and interpretation of the null and alternative hypotheses needs careful thought in the panel arena and second, there may be a problem of cross-sectional dependence in the errors across the unit root testing regressions. Some of the literature refers to the early studies that assumed cross-sectional independence as 'first generation' panel unit root tests, while the more recent approaches that allow for some form of dependence are termed 'second generation' tests.

A perhaps obvious starting point for unit root tests when one has a panel of data would be to run separate regressions over time for each series but to use Zellner's SUR approach, which we might term the multivariate ADF (MADF) test. This method can only be employed if $T >> N$, and Taylor and Sarno (1998) provide an early application to tests for purchasing power parity. However, it is fair to say that technique is now rarely used, researchers preferring instead to make use of the full panel structure.

A key consideration is the dimensions of the panel – is the situation that T is large or that N is large or both? If T is large and N small, the MADF approach can be used, although as Breitung and Pesaran (2008) note, in such a situation one may question whether it is worthwhile to adopt a panel approach at all, since for sufficiently large T, separate ADF tests ought to be reliable enough to render the panel approach hardly worth the additional complexity.

11.8.2 Tests with Common Alternative Hypotheses

Levin, Lin and Chu (2002) – hereafter LLC – develop a test based on the equation

$$\Delta y_{i,t} = \alpha_i + \theta_t + \delta_i t + \rho_i y_{i,t-1} + \sum \gamma_j \Delta y_{t-j} + v_{i,t} \qquad t = 1, 2, \ldots, T; i = 1, 2, \ldots, N \tag{11.18}$$

The model is very general since it allows for both entity-specific and time-specific effects through α_i and θ_t, respectively, as well as separate deterministic trends in each series through $\delta_i t$, and the lag structure to mop up autocorrelation in Δy. Of course, as for the Dickey–Fuller tests, any or all of these deterministic terms can be omitted from the regression. The null hypothesis is $H_0 : \rho_i \equiv \rho = 0 \ \forall \ i$ and the alternative is $H_1 : \rho < 0 \ \forall \ i$.

One of the reasons that unit root testing is more complex in the panel framework in practice is due to the plethora of 'nuisance parameters' in the equation which are necessary to allow for the fixed effects (i.e., the α_i, θ_t, $\delta_i t$). These nuisance parameters will affect the asymptotic distribution of the test statistics and hence LLC propose that two auxiliary regressions are run to remove their impacts. First, Δy_{it} is regressed on its lags, Δy_{it-j}, $j = 1, \ldots, p_i$ and on the exogenous variables (any or all from α_i, θ_t, and $\delta_i t$ as desired); the residuals, u_{1it} are obtained. Note that the numbers of lags of the dependent variables, p_i, need not be the same for each series in the panel. Next, the lagged level of y, y_{it-1}, is regressed on the same variables to get the residuals, u_{2it}. Then the residuals from both regressions are standardised by dividing them by the regression standard error, s_i, which is obtained from the augmented Dickey–Fuller regression in equation (11.18)

$$\tilde{u}_{1it} = u_{1it}/s_i \tag{11.19}$$

and

$$\tilde{u}_{2it} = u_{2it}/s_i \tag{11.20}$$

Thus $\tilde{u}_{1it}$ will be equivalent to Δy_{it} but with the effects of the deterministic components removed, and $\tilde{u}_{2it}$ will be equivalent to y_{it-1} but with the effects of the deterministic components removed. Finally, $\tilde{u}_{1it}$ is regressed on $\tilde{u}_{2it}$, and the slope estimate from this test regression is then used to construct a test statistic which is asymptotically distributed as a standard normal variate. The test statistic will approach this 'limiting' normal distribution as T tends to infinity and as N tends to infinity, although the convergence is faster for the former than the latter.

Breitung (2000) develops a modified version of the LLC test which does not include the deterministic terms (i.e., the fixed effects and/or a deterministic trend), and which standardises the residuals from the auxiliary regression in a more sophisticated fashion.

It should be clear that under the LLC and Breitung approaches, only evidence against the non-stationary null in one series is required before the joint null will be rejected. Breitung and Pesaran (2008) suggest that the appropriate conclusion when the null is rejected is that 'a significant proportion of the cross-sectional units are stationary'. Especially in the context of large N, this might not be very helpful since no information is provided on how many of the N series are stationary. Often, the homogeneity assumption is not economically meaningful either, since there is no theory suggesting that all of the series have the same autoregressive dynamics and thus the same value of ρ.

11.8.3 Panel Unit Root Tests with Heterogeneous Processes

The difficulty described at the end of the previous sub-section led Im, Pesaran and Shin (2003) – hereafter IPS – to propose an alternative approach where, given equation (11.18) as above, the null and alternative hypotheses are now $H_0 : \rho_i = 0 \ \forall \ i$ and $H_1 : \rho_i < 0, i = 1, 2, \ldots, N_1; \rho_i = 0, i = N_1 + 1, N_1 + 2, \ldots, N$.

So the null hypothesis still specifies all series in the panel as non-stationary, but under the alternative, a proportion of the series (N_1/N) are stationary, and the remaining proportion ($(N - N_1)/N$) are non-stationary. But it is clear that no restriction where all of the ρ are identical is imposed. The statistic for the panel test in this case is constructed by conducting separate unit root tests for each series in the panel, calculating the ADF t-statistic for each one in the standard fashion, and then taking their cross-sectional average. This average is then transformed into a standard normal variate under the null hypothesis of a unit root in all the series; IPS develop an LM-test approach as well as the more familiar t-test.[11] If the time series dimension is sufficiently large, it is then possible to run separate unit root tests on each series in order to determine the proportion for which the individual tests cause a rejection, and thus how strong is the weight of evidence against the joint null hypothesis.

It should be noted that while IPS's heterogeneous panel unit root tests are superior to the homogeneous case when N is modest relative to T, they may not be sufficiently powerful when N is large and T is small, in which case the LLC approach may be preferable.

Maddala and Wu (1999) and Choi (2001) developed a slight variant on the IPS approach based on an idea dating back to Fisher (1932), where unit root tests are again conducted separately on each series in the panel, and the p-values associated with the test statistics are then combined. If we call these p-values pv_i, $i = 1, 2, \ldots, N$, then under the null hypothesis of a unit root in each series, each pv_i will be distributed uniformly over the [0,1] interval and hence the following will hold for given N as $T \to \infty$

$$\lambda = -2\sum_{i=1}^{N} ln(pv_i) \sim \chi^2_{2N}. \quad \textbf{(11.21)}$$

The number of observations per series can differ in this case as the regressions are run separately for each series and then only their p-values are combined in the test statistic. Notice that the cross-sectional independence assumption is crucial here for this sum to follow a χ^2 distribution. Since the distribution of the ADF test statistic is non-standard and is dependent upon the inclusion of the nuisance parameters, unfortunately the p-values for inclusion in this equation must be obtained from a Monte Carlo simulation. Moreover, if the series under consideration have different lag lengths for Δy_{it} or there are different numbers of observations, each will require a separate Monte Carlo!

As well as the χ^2 statistic, Choi (2001) develops a variant of the test, still based on the p-values, that is asymptotically standard normally distributed. It should be evident that, like IPS, the Maddala–Wu–Choi approach does not require the same parameter, ρ, to apply to all of the series since the ADF test is run separately on each series in the panel.

11.8.4 Panel Stationarity Tests

The approaches described above are non-stationarity tests, and analogous to the Dickey–Fuller approach, they have non-stationarity under the null hypothesis. It is also possible, however, to construct a test where the null hypothesis is of stationarity for all series in the panel, analogous to the KPSS test of Kwaitkowski *et al.* (1992). In this case, the null hypothesis is that all of the series are stationary, which is rejected if at least one of them is non-stationary. This approach in the panel context was developed by

Hadri (2000), and leads to a test statistic that is asymptotically normally distributed. As in the univariate case, stationarity tests can be useful as a way to check for the robustness of the conclusions from unit root tests.

11.8.5 Allowing for Cross-Sectional Heterogeneity

The assumption of cross-sectional independence of the error terms in the panel regression is highly unrealistic and likely to be violated in practice. For example, in the context of testing for whether purchasing power parity holds, there are likely to be important unspecified factors that affect all exchange rates or groups of exchange rates in the sample, and will result in correlated residuals. O'Connell (1998) demonstrates the considerable size distortions that can arise when such cross-sectional dependencies are present but not accounted for – that is, the null hypothesis is rejected far too frequently when it is correct than should arise by chance alone if the distributional assumption holds for the test statistic. If the critical values employed in the tests are adjusted to remove the impacts of these size distortions, then the power of the tests will fall such that in extreme cases the benefit of using a panel structure could disappear completely. According to Maddala and Wu (1999), tests based on the Fisher statistic are more robust in the presence of unparameterised cross-sectional dependence than the IPS approach.

O'Connell proposes a feasible GLS estimator for ρ where an assumed non-zero form for the correlations between the disturbances is employed. To overcome the limitation that the correlation matrix must be specified (and this may be troublesome because it is not clear what form it should take), Bai and Ng (2004) propose an approach based on separating the data into a common factor component that is highly correlated across the series and a specific part that is idiosyncratic; a further approach is to proceed with OLS but to employ modified standard errors – so-called 'panel corrected standard errors' (PCSEs) – see, for example Breitung and Das (2005).

Overall, however, it is clear that satisfactorily dealing with cross-sectional dependence makes an already complex issue considerable harder still. In the presence of such dependencies, the test statistics are affected in a non-trivial way by the nuisance parameters. As a result, despite their inferiority in theory, the first generation approaches that ignore cross-sectional dependence are still widely employed in the empirical literature.

11.8.6 Panel Cointegration

It is often remarked in the literature that the development of the techniques for panel cointegration modelling is still in its infancy, while that for panel unit root testing is already quite mature. Testing for cointegration in panels is a rather complex issue, since one must consider the possibility of cointegration across groups of variables (what we might term 'cross-sectional cointegration') as well as within the groups. It is also possible that the parameters in the cointegrating series and even the number of cointegrating relationships could differ across the panel.

Most of the work so far has relied upon a generalisation of the single equation methods of the Engle–Granger type following the pioneering work by Pedroni (1999, 2004). His setup is very general and allows for separate intercepts for each group of potentially cointegrating variables and separate deterministic trends. For a set of variables y_{it} and $x_{m,i,t}$ that are individually integrated of order one and thought to be cointegrated

$$y_{it} = \alpha_i + \delta_i t + \beta_{1i} x_{1i,t} + \beta_{2i} x_{2i,t} + \ldots + \beta_{Mi} x_{Mi,t} + u_{i,t} \quad \textbf{(11.22)}$$

where $m = 1, \ldots, M$ are the explanatory variables in the potentially cointegrating regression; $t = 1, \ldots, T$ and $i = 1, \ldots, N$.

The residuals from this regression, $\hat{u}_{i,t}$ are then subjected to separate Dickey–Fuller or augmented Dickey–Fuller type regressions for each group of variables to determine whether they are I(1) – for example

$$\hat{u}_{i,t} = \rho_i \hat{u}_{i,t-1} + \sum_{j=1}^{p_i} \psi_{i,j} \Delta \hat{u}_{i,t-j} + v_{i,t} \quad \textbf{(11.23)}$$

The null hypothesis is that the residuals from all of the test regressions are unit root processes ($H_0 : \rho_i = 1$), and therefore that there is no cointegration. Pedroni proposes two possible alternative hypotheses – first, that all of the autoregressive dynamics are the same stationary process ($H_1 : \rho_i = \rho < 1 \ \forall \ i$) and second, that the dynamics from each test equation follow a different stationary process ($H_1 : \rho_i < 1 \ \forall \ i$). Hence, in the first case no heterogeneity is permitted, while in the second it is – analogous to the difference between LLC and IPS as described above. Pedroni then constructs a raft of different test statistics based on standardised versions of the usual t-ratio from equation (11.23). The standardisation required is a function of whether an intercept or trend is included in equation (11.23), and the value of M. These standardised test statistics are each asymptotically standard normally distributed.

Kao (1999) essentially develops a restricted version of Pedroni's approach, where the slope parameters in equation (11.22) are assumed to be fixed across the groups, although the intercepts are still permitted to vary. Then the DF or ADF test regression is run on a pooled sample assuming homogeneity in the value of ρ. These restrictions allow some simplification in the testing approach.

As well as testing for cointegration using the residuals following these extensions of Engle and Granger, it is also possible, although in general more complicated, to use a generalisation of the Johansen technique. This approach is deployed by Larsson, Lyhagen and Lothgren (2001), but a simpler alternative is to apply the Johansen approach to each group of series separately, collect the p-values for the trace test and then take −2 times the sum of their logs following Maddala and Wu (1999) as in equation (11.21) above. A full systems approach based on a 'global VAR' is possible but with considerable additional complexity – see Breitung and Pesaran (2008) and the many references therein for further details.

11.8.7 An Illustration of the Use of Panel unit Root and Cointegration Tests: The Link Between Financial Development and GDP Growth

An important issue for developing countries from a policy perspective is the extent to which economic growth and the sophistication of the country's financial markets are linked. It has been argued in the relevant literature that excessive government regulations (such as limits on lending, restrictions on lending and borrowing interest rates, the barring of foreign banks, etc.) may impede the development of the financial markets and consequently economic growth will be slower than if the financial markets were more vibrant. On the other hand, if economic agents are able to borrow at reasonable rates of interest or raise funding easily on the capital markets, this can increase the viability of real investment opportunities and allow for a more efficient allocation of capital.

Both the theoretical and empirical research in this area has led to mixed conclusions; the theoretical models arrive at different findings dependent upon the framework employed and the assumptions made. And on the empirical side, many existing studies in this area are beset by two issues: first, the direction of causality between economic and financial development could go the other way: if an economy grows, then the demand for financial products will itself increase. Thus it is possible that economic growth leads to financial market development rather than the other line of causality. Second, given that long time series are typically unavailable for developing economies, traditional unit root and cointegration tests that examine the link between these two variables suffer from low power. In particular, while research has been able to identify a link between economic growth and stock market development, such an effect could not be identified for the sophistication of the banking sector. This provides a strong motivation for the use of panel techniques, which are more powerful, and which constitute the approach adopted by Christopoulos and Tsionas (2004). Some of the key methodologies and findings of their paper will now be discussed.

Defining real output for country i as y_{it}, financial 'depth' as F, the proportion of total output that is investment as S, and the rate of inflation as p, the core model they employ is

$$y_{it} = \beta_{0i} + \beta_{1i}F_{it} + \beta_{2i}S_{it} + \beta_{3i}\dot{p}_{it} + u_{it}. \quad \textbf{(11.24)}$$

Financial depth, F, is proxied using the ratio of total bank liabilities to GDP. Christopoulos and Tsionas obtain data from the IMF's *International Financial Statistics* for ten countries (Colombia, Paraguay, Peru, Mexico, Ecuador, Honduras, Kenya, Thailand, the Dominican Republic and Jamaica) over the period 1970–2000.

The regression in equation (11.24) has national output as the dependent variable, and financial development as one of the independent variables, but Christopoulos and Tsionas also investigate the reverse causality with F as the dependent variable and y as one of the independent variables. They first apply unit root tests to each of the individual series (output, financial depth, investment share in GDP, and inflation) separately for the ten countries. The findings are mixed, but show that most series are best characterised by unit root processes in levels but are stationary in first differences. They then employ the panel unit root tests of Im, Pesaran and Shin, and the Maddala–Wu chi-squared test separately for each variable, but now using a panel comprising all ten countries. The number of lags of Δy_{it} is determined using AIC. The null hypothesis in all cases is that the process is a unit root. Now the results, presented here in Table 11.4, are much stronger and show conclusively that all four series are non-stationary in levels but stationary in differences.

Table 11.4 Panel unit root test results for economic growth and financial development

Variables	Levels		First differences	
	IPS	Maddala–Wu	IPS	Maddala–Wu
Output (y)	−0.18	27.12	−4.52***	58.33***
Financial depth (F)	2.71	14.77	−6.63***	83.64***
Investment share (S)	−0.04	30.37	−5.81***	62.98***
Inflation ($\dot{p}$)	−0.47	26.37	−5.19***	74.29***

Notes: The critical value for the Maddala–Wu test is 37.57 at the 1% level. *** denotes rejection of the null hypothesis of a unit root at the 1% level.

Source: Christopoulos and Tsionas (2004). Reprinted with the permission of Elsevier.

The next stage is to test whether the series are cointegrated, and again this is first conducted separately for each country and then using a panel approach. Focusing on the latter, the LLC approach is used along with the Harris–Tzavalis (1999) technique, which is broadly the same as LLC but has slightly different correction factors in the limiting distribution owing to its assumption that T is fixed as N tends to infinity. As discussed in the previous sub-section, these techniques are based on a unit root test on the residuals from the potentially cointegrating regression, and Christopoulos and Tsionis investigate the use of panel cointegration tests with fixed effects, and with both fixed effects and a deterministic trend in the test regressions. These are applied to the regressions both with y, and separately F, as the dependent variables.

The results in Table 11.5 quite strongly demonstrate that when the dependent variable is output, the LLC approach rejects the null hypothesis of a unit root in the potentially cointegrating regression residuals when fixed effects only are included in the test regression, but not when a trend is also included. In the context of the Harris–Tzavalis variant of the residuals-based test, for both the fixed effects and the fixed effects + trend regressions, the null is rejected. When financial depth is instead used as the dependent variable, none of these tests reject the null hypothesis. Thus, the weight of evidence from the residuals-based tests is that cointegration exists when output is the dependent variable, but it does not when financial depth is. The authors interpret this result as implying that causality runs from output to financial depth but not the other way around.

Table 11.5 Panel cointegration test results for economic growth and financial development

	LLC		Harris–Tzavalis	
	Fixed effects	**Fixed effects + trend**	**Fixed effects**	**Fixed effects + trend**
Dep. var.: y	-8.36^{***}	0.89	-77.13^{***}	-5.57^{***}
Dep. var.: F	−1.2 $r=0$	0.5 $r\leq1$	−0.85 $r\leq2$	−1.65 $r\leq3$
Fisher χ^2	76.09^{***}	30.73	28.91	23.26

Notes: 'Dep. var.' denotes the dependent variable; *** denotes rejection of the null hypothesis of no cointegration at the 2% level. The critical values for the Fisher test are 37.57 and 31.41 at the 1% and 5% levels, respectively.

Source: Christopoulos and Tsionas (2004). Reprinted with the permission of Elsevier.

In the final row of Table 11.5, a systems approach to testing for cointegration, based on the sum of the logs of the *p*-values from the Johansen test, shows that the null hypothesis of no cointegrating vectors ($H_0 : r = 0$) is rejected, while ($H_0 : r \leq 1$) and above are all not rejected. Thus the conclusion is that one cointegrating relationship exists between the four variables across the panel. Note that in this case, since cointegration is tested within a VAR system, all variables are treated in parallel, and hence there are not separate results for different dependent variables.

11.9 Further Feading

Some readers may feel that further instruction in this area could be useful. If so, the classic specialist references to panel data techniques are Baltagi (2005) and Hsiao (2003) and further references are Arellano (2003) and Wooldridge (2010). All four are extremely detailed and have excellent referencing to recent developments in the theory of panel model specification, estimation and testing. However, all also require a high level of mathematical and econometric ability on the part of the reader. A more intuitive and accessible, but less detailed, treatment is given in Kennedy (2003, Chapter 17). Some examples of financial studies that employ panel techniques and outline the methodology sufficiently descriptively to be worth reading as aids to learning are given in the examples above. The book by Maddala and Kim (1999) provides a fairly accessible treatment of unit roots and cointegration generally, although the time of publication implies that the most recent developments are excluded. Breitung and Pesaran (2008) is a more recent survey and is comprehensive, but at a higher technical level.

KEY CONCEPTS

The key terms to be able to define and explain from this chapter are

- pooled data
- fixed effects
- random effects
- within transform
- between estimation
- panel cointegration test
- seemingly unrelated regression
- least squares dummy variable estimation
- Hausman test
- time-fixed effects
- panel unit root test

SELF-STUDY QUESTIONS

1. (a) What are the advantages of constructing a panel of data, if one is available, rather than using pooled data?

(b) What is meant by the term 'seemingly unrelated regression'? Give examples from finance of where such an approach may be used.

(c) Distinguish between balanced and unbalanced panels, giving examples of each.

2. (a) Explain how fixed effects models are equivalent to an ordinary least squares regression with dummy variables.

(b) How does the random effects model capture cross-sectional heterogeneity in the intercept term?

(c) What are the relative advantages and disadvantages of the fixed versus random effects specifications and how would you choose between them for application to a particular problem?

3. Find a further example of where panel regression models have been used in the academic finance literature and do the following:

- Explain why the panel approach was used.
- Was a fixed effects or random effects model chosen and why?
- What were the main results of the study and is any indication given about whether the results would have been different had a pooled regression been employed instead in this or in previous studies?

4. (a) What are the advantages and disadvantages of conducting unit root tests within a panel framework rather than series by series?

(b) Explain the differences between panel unit root tests based on a common alternative hypothesis and those based on heterogeneous processes.

[1] Hence, strictly, if the data are not on the same entities (for example, different firms or people) measured over time, then this would not be panel data.

[2] Note that *k* is defined slightly differently in this chapter compared with others in the book. Here, *k* represents the number of slope parameters to be estimated (rather than the total number of parameters as it is elsewhere), which is equal to the number of explanatory variables in the regression model.

[3] For example, the SUR framework has been used to test the impact of the introduction of the euro on the integration of European stock markets (Kim, Moshirian and Wu, 2005), in tests of the CAPM, and in tests of the forward rate unbiasedness hypothesis (Hodgson, Linton and Vorkink, 2004).

[4] It is important to recognise this limitation of panel data techniques that the relationship between the explained and explanatory variables is assumed constant both cross-sectionally and over time, even if the varying intercepts allow the average values to differ. The use of panel techniques rather than estimating separate time series regressions for each object or estimating separate cross-sectional regressions for each time period thus implicitly assumes that the efficiency gains from doing so outweigh any biases that

may arise in the parameter estimation.

[5] It is known as the *within transformation* because the subtraction is made within each cross-sectional object.

[6] An advantage of running the regression on average values (the *between estimator*) over running it on the demeaned values (the *within estimator*) is that the process of averaging is likely to reduce the effect of measurement error in the variables on the estimation process.

[7] Interestingly, while many casual observers believe that concentration in UK retail banking has grown considerably, it actually fell slightly between 1986 and 2002.

[8] A Chow test for structural stability reveals a structural break between the two sub-samples. No other commentary on the results of the equilibrium regression is given by the authors.

[9] The notation used here is a slightly modified version of Kennedy (2003, p. 315).

[10] de Haas and van Lelyveld employ corrections to the standard errors for heteroscedasticity and autocorrelation. They additionally conduct regressions including interactive dummy variables, although these are not discussed here.

[11] Both tests presume that there is a balanced panel – that is, the number of time series observations is the same for each cross-sectional entity.

12

Limited Dependent Variable Models

LEARNING OUTCOMES

In this chapter, you will learn how to

- Compare between different types of limited dependent variables and select the appropriate model
- Interpret and evaluate logit and probit models
- Distinguish between the binomial and multinomial cases
- Deal appropriately with censored and truncated dependent variables

12.1 Introduction and Motivation

Chapters 5 and 10 have shown various uses of dummy variables to numerically capture the information qualitative variables – for example, day-of-the-week effects, gender, credit ratings, etc. When a dummy is used as an explanatory variable in a regression model, this usually does not give rise to any particular problems (so long as one is careful to avoid the *dummy variable trap* – see Chapter 10). However, there are many situations in financial research where it is the explained variable, rather than one or more of the explanatory variables, that is qualitative. The qualitative information would then be coded as a dummy variable and the situation would be referred to as a *idiscrete choice* variable and needs to be treated differently. The term refers to any problem where the values that the dependent variables may take are limited to certain integers (e.g., 0, 1, 2, 3, 4) or even where it is a binary number (only 0 or 1, which would then be known as a *binary choice* variable).

Discrete choice variables are one set from among what are known more generally as *limited dependent variables*, since the values they can take are limited to only certain integers. Another class of limited dependent variables are where the data that we see are censored or truncated in some way – in other words, we can only observe the true values for part of the distribution while for the remainder above or below some fixed threshold, the true values remain latent. We will return to censored and truncated series – and the differences between them – later in the chapter.

There are numerous examples of instances where the dependent variable may arise from a binary choice, 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 in Section 12.9. We will first examine a simple and obvious, but unfortunately flawed, method for dealing with binary dependent variables, known as the *linear probability model*.

12.2 The Linear Probability Model

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

$$P_i = p(y_i = 1) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_k x_{ki} + u_i, \quad i = 1, \ldots, N \tag{12.1}$$

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 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 ($y_i = 1$) as a function of its market capitalisation (x_{2i}, measured in millions of US dollars), and we fit the following line:

$$\hat{P}_i = -0.3 + 0.012x_{2i} \tag{12.2}$$

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 $-0.3 + 0.012 \times 50 = 0.3$ (or 30%) probability of making a dividend payment. Graphically, this situation may be represented as in Figure 12.1.

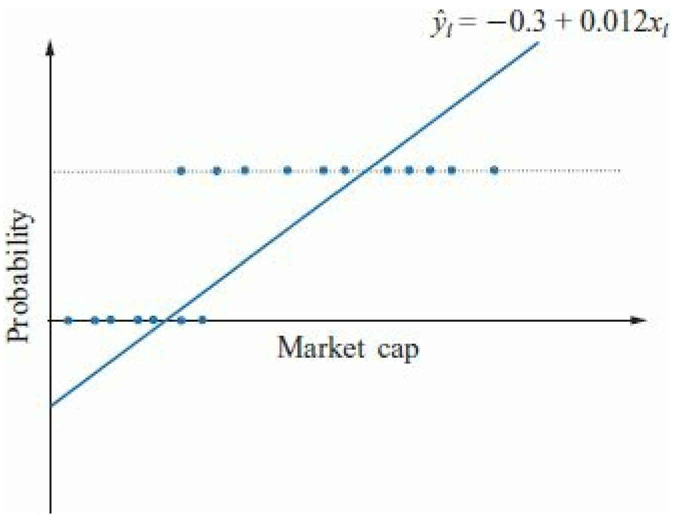

Figure 12.1 The fatal flaw of the linear probability model

While the linear probability model is simple to estimate and intuitive to interpret, the diagram 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. However, there are at least two reasons why this is still not adequate

(1) The process of truncation will result in too many observations for which the estimated probabilities are exactly zero or one.

(2) 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, so a different kind of model is usually used for binary dependent variables – either a *logit* or a *probit* specification. These approaches will be discussed in the following sections. But before moving on, it is worth noting that the LPM also suffers from a couple of more standard econometric problems that we have examined in previous chapters. First, since the dependent variable takes only one of two values, for given (fixed in repeated samples) values of the explanatory variables, the disturbance term will also take on only one of two values.[1] Consider again equation (12.1). If $y_i = 1$, then by definition

$$u_i = 1 - \beta_1 - \beta_2 x_{2i} - \beta_3 x_{3i} - \cdots - \beta_k x_{ki};$$

but if $y_i = 0$, then

$$u_i = -\beta_1 - \beta_2 x_{2i} - \beta_3 x_{3i} - \cdots - \beta_k x_{ki}.$$

Hence the error term cannot plausibly be assumed to be normally distributed. Since u_i changes systematically with the explanatory variables, the disturbances will also be heteroscedastic. It is therefore essential that heteroscedasticity-robust standard errors are always used in the context of limited dependent variable models.

12.3 The Logit Model

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. This is shown in Figure 12.2.

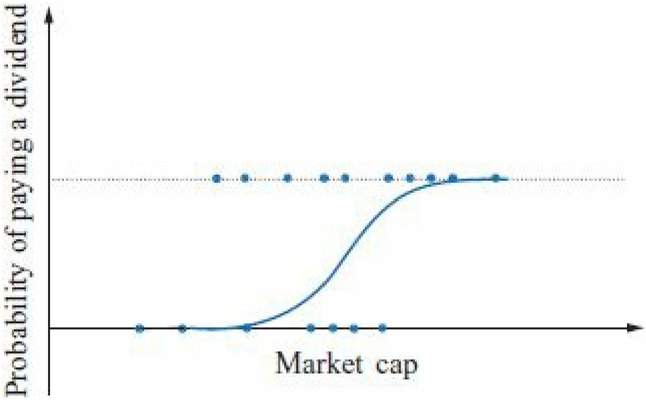

Figure 12.2 The logit model

The logistic function F, which is a function of any random variable, z, would be

$$F(z_i) = \frac{e^{z_i}}{1 + e^{z_i}} = \frac{1}{1 + e^{-z_i}} \tag{12.3}$$

where e is the exponential under the logit approach. The model is so called because the function F is in fact the cumulative logistic distribution. So the logistic model estimated would be

$$P_i = \frac{1}{1 + e^{-(\beta_1 + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} + u_i)}} \tag{12.4}$$

where again P_i is the probability that $y_i = 1$.

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. In equation (12.3), as z_i tends to infinity, e^{-z_i} tends to zero and $1/(1 + e^{-z_i})$ tends to 1; as z_i tends to minus infinity, e^{-z_i} tends to infinity and $1/(1 + e^{-z_i})$ tends to 0.

Clearly, this 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 – this is discussed in Section 12.7 and in more detail in the appendix to this chapter.

12.4 Using a Logit to Test the Pecking Order

Hypothesis

This section examines a study of the pecking order hypothesis due to Helwege and Liang (1996). 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', initially proposed by Myers (1984). 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.[2] 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.

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 1984–92 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) 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. As such, the dependent variable will be binary – that is, a column of 1s (firm raises funds externally) and 0s (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. Hence 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 1983–92 period; 'previous financing' is a dummy variable equal to 1 for firms that obtained external financing in the previous year. The results from the logit regression are presented in Table 12.1.

Table 12.1 Logit estimation of the probability of external financing

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 × operating income			−0.03 (−0.59)
Assets	0.0004 (1.99)	0.0003 (1.36)	0.0004 (1.99)
Industry asset growth	−0.002 (−1.70)	−0.002 (−1.35)	−0.002 (−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.

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.[3] 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.

12.5 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. The function F in equation (12.3) is replaced by

$$F(z_i) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_i} e^{-\frac{z_i^2}{2}} dz \quad \textbf{(12.5)}$$

This function is the cumulative distribution function for a standard normally distributed random variable. As for the logistic approach, this function provides a transformation to ensure that the fitted probabilities will lie between zero and one. Also as for the logit model, the marginal impact of a unit change in an explanatory variable, x_{4i} say, will be given by $\beta_4 F(z_i)$, where β_4 is the parameter attached to x_{4i} and $z_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + u_i$.

12.6 Choosing Between the Logit and Probit Models

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 (such as figure 12.2) 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 (2011) 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.

12.7 Estimation of Limited Dependent Variable Models

Given that both logit and probit are non-linear models, they cannot be estimated by OLS. While the parameters could, in principle, be estimated using non-linear least squares (NLS), maximum likelihood (ML) is simpler and is invariably used in practice. As discussed in Chapter 9, 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 the logit or probit model is used, but the general principles for parameter estimation described in Chapter 9 will still apply. That is, we form the appropriate log-likelihood function and then the software package will find the values of the parameters that jointly maximise it using an iterative search procedure. A derivation of the ML estimator for logit and probit models is given in the appendix to this chapter. Box 12.1 shows how to interpret the estimated parameters from probit and logit models.

BOX 12.1 Parameter interpretation for probit and logit 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 $100 \times \beta_2$% 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 models, this interpretation would be incorrect because the form of the function is not $P_i = \beta_i + \beta_2 x_i + u_i$, for example, but rather $P_i = F(\beta_i + \beta_2 x_i + u_i)$, where F represents the (non-linear) logistic function. 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 $F(x_{2i})(1 - F(x_{2i}))$. So in fact, a 1-unit increase in x_{2i} will cause a $\beta_2 F(x_{2i})(1 - F(x_{2i}))$ increase in probability. Usually, these impacts of incremental changes in an explanatory variable are evaluated by setting each of them to their mean values. For example, suppose we have estimated the following logit model with three explanatory variables using maximum likelihood

$$\hat{P}_i = \frac{1}{1 + e^{-(0.1 + 0.3x_{2i} - 0.6x_{3i} + 0.9x_{4i})}} \quad \textbf{(12.7)}$$

Thus we have $\hat{\beta}_1 = 0.1$, $\hat{\beta}_2 = 0.3$, $\hat{\beta}_3 = -0.6$, $\hat{\beta}_4 = 0.9$. We now need to