FINANCIAL
ECONOMETRICS

Peijie Wang

Financial Econometrics

Set against a backdrop of rapid expansions of interest in the modelling and analysis of financial data and the problems to which they are applied, this textbook, now in its second edition, offers an overview and discussion of the contemporary topics surrounding financial econometrics, covering all major developments in the area in recent years in an informative and succinct way.

Extended from the first edition of mainly time series modelling, the new edition also takes in discrete choice models, estimation of censored and truncated samples, as well as panel data analysis that has witnessed phenomenal expansion in application in finance and financial economics since the publication of the first edition of the book. Virtually all major topics on time series, cross-sectional and panel data analysis have been dealt with. Subjects covered include:

- unit roots, cointegration and other comovements in time series
- time varying volatility models of the GARCH type and the stochastic volatility approach
- analysis of shock persistence and impulse responses
- Markov switching
- present value relations and data characteristics
- state space models and the Kalman filter
- frequency domain analysis of time series
- limited dependent variables and discrete choice models
- truncated and censored samples
- panel data analysis

Refreshingly, every chapter has a section of two or more examples and a section of empirical literature, offering the reader the opportunity to practise right away the kind of research going on in the area. This approach helps the reader develop interest, confidence and momentum in learning contemporary econometric topics.

Graduate and advanced undergraduate students requiring a broad knowledge of techniques applied in the finance literature, as well as students of financial economics engaged in empirical enquiry, should find this textbook to be invaluable.

Peijie Wang is Professor of Finance at IÉSEG School of Management, Catholic University of Lille. He is author of *An Econometric Analysis of the Real Estate Market* (Routledge 2001) and *The Economics of Foreign Exchange and Global Finance*.

Routledge Advanced Texts in Economics and Finance

Financial Econometrics

Second edition

Peijie Wang

Routledge
Taylor & Francis Group

Contents

Figures

Tables

Acknowledgements

the quality and appeal of the first edition that made the book evolve into a new edition.

Finally, I thank the production and marketing teams of Routledge who bring the book to the reader.

PJW
January 2008

Preface

This book focuses on econometric models widely and frequently used in the examination of issues in financial economics and financial markets, which are scattered in the literature yet to be integrated into a single-volume, multi-theme, and empirical research-oriented text. The book, providing an overview of contemporary topics related to the modelling and analysis of financial data, is set against a backdrop of rapid expansions of interest in both the models themselves and the financial problems to which they are applied. Extended from the first edition of mainly time series modelling, the new edition also takes in discrete choice models, estimation of censored and truncated samples, as well as panel data analysis that has witnessed phenomenal expansion in application in finance and financial economics since the publication of the first edition of the book. Virtually all major topics on time series, cross-sectional and panel data analysis have been dealt with.

We assume that the reader has already had knowledge in econometrics and finance at the intermediate level. So basic regression analysis and time series models such as the OLS, maximum likelihood and ARIMA, while being referred to from time to time in the book, are only briefly reviewed but are not brought up as a book topic; nor the concept of market efficiency and models for asset pricing. For the former, there are good books such as *Basic Econometrics* by Gujarati (2002), *Econometric Analysis* by Greene (2008), and *Introduction to Econometrics* by Maddala (2001); and for the latter, the reader is recommended to refer to *Principles of Corporate Finance* by Brealey, Myers and Allen (2006), *Corporate Finance* by Ross, Westerfield and Jaffe (2008), *Investments* by Sharpe, Alexander and Bailey (1999), *Investments* by Bodie, Kane and Marcus, (2008), and *Financial Markets and Corporate Strategy* by Grinblatt and Titman (2002).

The book has two unique features – every chapter (except the first three introductory chapters and the final chapter) has a section of two or more examples and cases, and a section of empirical literature, offering the reader the opportunity to practice right away the kind of research in the area. The examples and cases, either from the literature or of the book itself, are well executed, and the results are explained in detail in plain language. This would, as we hope, help the reader gain interest, confidence, and momentum in learning contemporary econometric topics. At the same time, the reader would find that the way of implementation

and estimation of a model is unavoidably influenced by the view of the researcher on the issue in a social science subject; nevertheless, for a serious researcher, it is not easy to make two plus two equal to any desired number she or he wants to get. The empirical literature reviewed in each chapter is comprehensive and up to date, exemplifying rich application areas at both macro and micro levels limited only by the imagination of human beings. The section demonstrates how a model can and should match practical problems coherently and guide the researcher's consideration on the rationale, methodology and factors in the research. Overall, the book is methods, models, theories, procedures, surveys, thoughts and tools.

To further help the reader carry out an empirical modern financial econometrics project, the book introduces research tools and sources of information in the final chapter. These include on-line information on, and the websites for, the literature on research in financial economics and financial markets; commonly used econometric software packages for time series, cross-sectional and panel data analysis; professional associations and learned societies; and international and national institutions and organisations. A website link is provided whenever it is possible. The provision is based on our belief that, to perfect an empirical study, one has to understand the wider background of the business environment, market operations and institutional roles, and to frequently upgrade and update the knowledge base which is nowadays largely through internet links.

The book can be used in graduate programmes in financial economics, financial econometrics, international finance, banking and investment. It can also be used as doctorate research methodology materials and by individual researchers interested in econometric modelling, financial studies, or policy evaluation.

References

Bodie, Zvi, Kane, Alex and Marcus, Alan, J. (2008), *Investments* 7th edn, McGraw-Hill.

Brealey, Richard, A., Myers, Stewart, C. and Allen, Franklin (2006), *Principles of Corporate Finance* 8th ed, McGraw-Hill.

Greene, William, H. (2008), *Econometric Analysis* 8th edn, Prentice Hall.

Grinblatt, Mark and Titman, Sheridan (2002), *Financial Markets and Corporate Strategy* 2nd ed, McGraw-Hill.

Gujarati, Damodar (2002), *Basic Econometrics* 4th edn, McGraw-Hill.

Maddala, G.S. (2001), *Introduction to Econometrics* 3rd edn, Wiley.

Ross, Stephen, A., Westerfield, Randolph, W. and Jaffe, Jeffrey (2008), *Corporate Finance* 8th edn, McGraw-Hill.

Sharpe, William, F., Alexander, Gordon, J. and Bailey, Jeffery, V. (1999), *Investments* 6th edn, Prentice-Hall International.

1 Stochastic processes and financial data generating processes

1.1. Introduction

Statistics is the analysis of events and the association of events, with a probability. Econometrics pays attention to economic events, the association between these events, and between these events and human beings' decision-making – government policy, firms' financial leverage, individuals' investment/consumption choice, and so on. The topics of this book, financial econometrics, focus on the variables and issues of financial economics, the financial market and the participants.

The financial world is an uncertain universe where events take place every day, every hour, and every second. Information arrives randomly and so do the events. Nonetheless, there are regularities and patterns in the variables to be identified, effect of a change on the variables to be assessed, and links between the variables to be established. Financial econometrics attempts to perform the analysis of these kinds through employing and developing various relevant statistical procedures.

There are generally three types of economic and financial variables – the rate variable, the level variable and the ratio variable. The first category measures the speed at which, for example, wealth is generated, or savings are made, at one point of time (continuous time) or in a short interval of time (discrete time). The rate of return on a company's stock or share is a typical rate variable. The second category works out the amount of wealth, such as income and assets, being accumulated over a period (continuous time) or in a few of short time intervals (discrete time). A firm's assets and a country's GDP are typical level variables, though they are different in a certain sense in that the former is a stock variable and the latter is a flow variable. The third category consists of two sub-categories, one is the type I ratio variable or the component ratio variable, and the other is the type II ratio, the contemporaneous relativity ratio variable. The unemployment rate is rather a ratio variable, a type I ratio variable, than a rate variable. The exchange rate is more precisely a typical type II ratio variable instead of a rate variable. This classification of variables does not necessarily correspond to the classification of variables into flow variables and stock variables in economics. For example, we will see in Chapter 8 that both income and value should behave similarly in terms

of statistical characteristics as non-stationary variables, though the former is a flow variable and the latter a stock variable, if the fundamental relationship between them is to hold.

Before we can establish links and chains of influence amongst the variables in concern, which are in general random or stochastic, we have to assess their individual characteristics first. That is, with what probability may the variable take a certain value, or how likely may an event (the variable taking a given value) occur? Such assessment of the characteristics of individual variables is made through the analysis of their statistical distributions. Bearing this in mind, a number of stochastic processes, which are commonly encountered in empirical research in economics and finance, are presented, compared and summarised in the next section. The behaviour and valuation of economic and financial variables are discussed in association with these stochastic processes in Section 1.3, with further extension and generalisation.

Independent identical distribution (iid) and normality in statistical distributions are commonly supposed to be met, though from time to time we would modify the assumptions to fit the real world problem more appropriately. If the rate variables are, as widely assumed, iid and normally distributed around a constant mean, then its corresponding level variable would be log normally distributed around a mean which is increasing exponentially over time, and the level variable in logarithms is normally distributed around a mean which is increasing linearly over time. This is the reason why we usually work with the level variables in their logarithms.

Prior to proceeding to the main topics of this book, a few of most commonly assumed statistical distributions applied in various subjects are reviewed in Chapter 2, in conjunction with their rationale in statistics and relevance in finance. The examination of statistical distributions of stochastic variables helps assess their characteristics and understand their behaviour. Then, primary statistical estimation methods, covering the ordinary least squares, the maximum likelihood method and the method of moments and the generalised method of moment, are briefly reviewed in Chapter 3. The iid and iid under normal distributions are firstly assumed for the residuals from fitting the model. Then, the iid requirements are gradually relaxed, leading to general residual distributions typically observed in time series and cross-section modelling. This also serves the purpose of introducing elementary time series and cross-section models and specifications, based on which and from which most models in the following chapters of this book are developed and evolved.

The classification of financial variables into rate variables and level variables gives rise to stationarity and non-stationarity in financial time series, though there might be no clear-cut match of the economic and financial characteristic and the statistical characteristic in empirical research; whilst the behaviour and properties of ratio variables may be even more controversial. Related to this issue, Chapter 4 analyses unit roots and presents the procedures for testing for unit roots. Then the chapter introduces the idea of cointegration where a combination of two or more non-stationary variables becomes stationary. This is a special type of link amongst stochastic variables, implying that there exists a so-called long-run relationship.

The chapter also extends the analysis to cover common trends and common cycles, the other major types of links amongst stochastic variables in economics and finance.

One of the violations to the iid assumption is heteroscedasticity, i.e. the variance is not the same from each of the residuals; and modifications are consequently required in the estimation procedure. The basics of this issue and the ways to handle it have been learned from introductory econometrics or statistics or can be learned in Chapter 3 on overview of estimation methods. What we introduce here in Chapter 5 is specifically a kind of variance which changes with time, or time-varying variance. Time-varying variance or time-varying volatility is frequently found in many financial time series so has to be dealt with seriously. Two types of time-varying volatility models are discussed, one is GARCH (Generalised Auto Regressive Conditional Heteroscedasticity) and the other is stochastic volatility.

How persistent is the effect of a shock is important in financial markets. It is not only related to the response of, say, financial markets to a piece of news, but is also related to policy changes, of the government or the firm. This issue is addressed in Chapter 6, which also incorporates impulse response analysis, a related subject which we reckon should be under the same umbrella. Regime shifts are important in the economy and financial markets as well, in that regime shifts or breaks in the economy and market conditions are often observed, but the difficulties are that regime shifts are not easily captured by conventional regressional analysis and modelling. Therefore, Markov switching is introduced in Chapter 7 to handle the issues more effectively. The approach helps improve our understanding about an economic process and its evolving mechanism constructively.

Some economic and financial variables have built-in fundamental relationships between them. One of such fundamental relationships is that between income and value. Economists regard that the value of an asset is derived from its future income generating power. The higher the income generating power, the more valuable is the asset. Nevertheless, whether this law governing the relationship between income and value holds is subject to empirical scrutiny. Chapter 8 addresses this issue with the help of econometric procedures, which identify and examine the time series characteristics of the variables involved.

Econometric analysis can be carried out in the conventional time domain as discussed in the above, and can also be performed through some transformations. Analysis in the state space is one of such endeavours, presented in Chapter 9. What the state space does is to model the underlying mechanisms through the changes and transitions in the state of its unobserved components, and establish the links between the variables of concern, which are observed, and those unobserved state variables. It explains the behaviour of externally observed variables by examining the internal, dynamic and systematic changes and transitions of unobserved state variables, to reveal the nature and cause of the dynamic movement of the variables effectively. State space analysis is usually executed with the help of the Kalman filter, also introduced in the chapter.

State space analysis is nonetheless still in the time domain, though it is not the conventional time domain analysis. With spectral analysis of time series in

Chapter 10, estimation is performed in the frequency domain. That is, time domain variables are transformed into frequency domain variables prior to the analysis, and the results in the frequency domain may be transformed back to the time domain when necessary. Such transformations are usually achieved through the Fourier transform and the inverse Fourier transform and, in practice, through the Fast Fourier Transform (FFT) and the Inverse Fast Fourier Transform (IFFT). The frequency domain properties of variables are featured by their spectrum, phase and coherence, to reflect individual time series' characteristics and the association between several time series, in the ways similar to those in the time domain.

Deviating from the preceding chapters of the book, Chapters 11 and 12 study models with limited dependent variables. The dependent variable in these two chapters is not observed on the whole range or whole sample, it is discrete, censored or truncated. Also deviating from the preceding chapters, data sets analysed in Chapters 11 and 12 are primarily cross-sectional. That is, they are data for multiple entities, such as individuals, firms, regions or countries, considered to be observed at a single time point. Issues associated with choice are addressed in Chapter 11. Firms and individuals encounter choice problems from time to time. Choice is deeply associated with people's daily life, firms' financing and investment activities, managers' business dealings and financial market operations. In financial terms, people make decisions on choice aimed at achieving higher utility of their work, consumption, savings, investment and their combinations. Firms make investment, financing and other decisions, supposedly aimed at maximising shareholder value. A firm may choose to use financial derivatives to hedge interest rate risk, or choose not to use financial derivatives. A firm may decide to expand its business into foreign markets, or not to expand into foreign markets. The above choice problems can be featured by binary choice models where the number of alternatives or options is two, in a usual frame of 'to do' or 'not to do'. General discrete choice models emerge when the number of alternatives or options is extended to be more than two. Since discrete choice models are non-linear, marginal effects are specifically considered.

In addition to discrete choice models where a dependent variable possesses discrete values, the values of dependent variables can also be censored or truncated. Chapter 12 examines issues in estimation of models involving limited dependent variables with regard to censored and truncated samples. Estimation of truncated or censored samples with certain conventional regression procedures can cause bias in parameter estimation. This chapter discusses the causes of the bias and introduces pertinent procedures to correct the bias arising from truncation and censoring, as well as the estimation procedures that produce unbiased parameter estimates. A wider issue of selection bias is specifically addressed.

The use of panel data and application of panel data modelling have increased drastically in the last five years in finance and related areas. The volume of studies and papers employing panel data has been multifold, in recognition of the advantages offered by panel data approaches as well as panel data sets themselves, and in response to the growing availability of data sets in the form of panel. Chapter 13 introduces various panel data models and model specifications and addresses various issues in panel data model estimation. Panel data covered in this

chapter refer to data sets consisting of cross-sectional observations over time, or pooled cross-section and time series data. They have two dimensions, one for time and one for the cross-section entity. Two major features that do not exist with the one-dimension time series data or the one-dimension cross-sectional data are fixed effects and random effects, which are analysed, along with the estimation of fixed effects models, random effects models and random parameter models. Issues of bias in parameter estimation for dynamic panel data models are then addressed and a few approaches to estimating dynamic panel models are presented.

Financial econometrics is only made possible by the availability of vast economic and financial data. Problems and issues in the real world have inspired the generation of new ideas and stimulated the development of more powerful procedures. The last chapter of the book, Chapter 14, is written to make such a real world and working environment immediately accessible by the researcher, providing information on the sources of literature and data, econometric software packages and organisations and institutions ranging from learned societies and regulators to market players.

1.2. Stochastic processes and their properties

The rest of this chapter presents stochastic processes frequently found in the financial economics literature and relevant to such important studies as market efficiency and rationality. In addition, a few terms fundamental to modelling financial time series are introduced. The chapter discusses stochastic processes in the tradition of mathematical finance, as we feel that there rarely exist links, at least explicitly, between mathematical finance and financial econometrics, to demonstrate the rich statistical properties of financial securities and their economic rationale ultimately underpinning the evolution of the stochastic process. After providing definitions and brief discussions of elementary stochastic processes in the next section, we begin with the generalisation of the Wiener process in Section 1.3, and gradually progress to show that the time path of many financial securities can be described by the Wiener process and its generalisations which can accommodate such well known econometric models or issues as ARIMA (Auto Regressive Integrated Moving Average), GARCH (Generalised Auto Regressive Conditional Heteroscedasticity), stochastic volatility, stationarity, mean-reversion, error correction and so on. Throughout the chapter, we do not particularly distinguish discrete and continuous time series and what matters to the analysis is that the time interval is small enough. The results are almost identical though this treatment does provide more intuition to real world problems. There are many stochastic processes books available, e.g., Ross (1996) and Medhi (1982). For modelling of financial securities, interested readers can refer to Jarrow and Turnbull (1999).

1.2.1. Martingales

A stochastic process X_n $(n = 1, 2, \ldots)$, with $E\left[X_n\right] < \infty$ for all n, is a martingale, if:

$$E\left[X_{n+1} \mid X_1, \ldots X_n\right] = X_n \tag{1.1}$$

Further, if a stochastic process X_n $(n = 1, 2, \ldots)$, with $E\left[X_n\right] < \infty$ for all n, is a submartingale, if:

$$E\left[X_{n+1} \mid X_1, \ldots X_n\right] \geq X_n \tag{1.2}$$

and is a supermartingale if:

$$E\left[X_{n+1} \mid X_1, \ldots X_n\right] \leq X_n \tag{1.3}$$

1.2.2. Random walks

A random walk is the sum of a sequence of independent and identically distributed (iid) variables X_i $(i = 1, 2, \ldots)$, with $E\left[X_i\right] < \infty$. Define:

$$S_n = \sum_{i=1}^{n} X_i \tag{1.4}$$

S_n is referred as a random walk. When X_i takes only two values, $+1$ and -1, with $P\{X_i = 1\} = p$ and $P\{X_i = -1\} = 1 - p$, the process is named as the Bernoulli random walk. If $p = 1 - p = \frac{1}{2}$, the process is called a simple random walk.

1.2.3. Gaussian white noise processes

A Gaussian process, or Gaussian white noise process, or simply white noise process, X_n, $(n = 1, 2, \ldots)$ is a sequence of independent random variables, each of which has a normal distribution:

$$X_n \sim N(0, \sigma^2) \tag{1.5}$$

with the probability density function being:

$$f_n(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-(x^2/2\sigma^2)} \tag{1.6}$$

The sequence of these independent random variables of the Gaussian white noise has a multivariate normal distribution and the covariance between any two variables in the sequence, $\text{Cov}(X_j, X_k) = 0$ for all $j \neq k$.

A Gaussian process is a white noise process because, in the frequency domain, it has equal magnitude in every frequency, or equal component in every colour. We know that the light with equal colour components, such as sunlight, is white. Readers interested in frequency domain analysis can refer to Chapter 10 for details.

1.2.4. Poisson processes

A Poisson process $N(t)$ $(t \geq 0)$ is a counting process where $N(t)$ is an integer representing the number of 'events' that have occurred up to time t, and the process

has independent increments, i.e. the number of events have occurred in interval $(s, t]$ is independent from the number of events in interval $(s + \tau, t + \tau]$.

Poisson processes can be stationary and non-stationary. A stationary Poisson process has stationary increments, i.e. the probability distribution of the number of events occurred in any interval of time is only dependent on the length of the time interval:

$$P\{N(t + \tau) - N(s + \tau)\} = P\{N(t) - N(s)\} \tag{1.7}$$

Then the probability distribution of the number of events in any time length τ is:

$$P\{N(t + \tau) - N(t) = n\} = e^{-\lambda t} \frac{(\lambda t)^n}{n!} \tag{1.8}$$

where λ is called the arrival rate, or simply the rate of the process. It can be shown that:

$$E\{N(t)\} = \lambda t, \qquad \mathrm{Var}\{N(t)\} = \lambda t \tag{1.9}$$

In the case that a Poisson process is non-stationary, the arrival rate is a function of time, thereby the process does not have a constant mean and variance.

1.2.5. Markov processes

A sequence X_n $(n = 0, 1, \ldots)$ is a Markov process if it has the following property:

$$P\{X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, X_1 = x_1, X_0 = x_0\} = P\{X_{n+1} = x_n \mid X_n = x_n\} \tag{1.10}$$

The Bernoulli random walk and simple random walk are the cases of Markov processes. It can be shown that the Poisson process is a Markov process as well.

A discrete time Markov process that takes finite or countable number of integer values x_n, is called a Markov chain.

1.2.6. Wiener processes

A Wiener process, also known as Brownian motion, is indeed the very basic element in stochastic processes:

$$\left. \begin{aligned} \Delta z(t) &= \varepsilon \sqrt{\Delta t}, \qquad \Delta t \to 0 \\ \varepsilon &\sim N(0, 1) \end{aligned} \right\} \tag{1.11}$$

The Wiener process can be derived from the simple random walk, replacing time sequence by time series when time intervals become smaller and smaller and

approach zero. If $z(t)$ is a simple random walk such that it moves forward and backward by Δz in time interval Δt, then:

$$E\left[z(t)\right] = 0$$

$$\mathrm{Var}\left[z(t)\right] = (\Delta z)^2 \frac{t}{\Delta t} \tag{1.12}$$

In a sensible and convenient way, let the distance of the small move $\Delta z = \sqrt{\Delta t}$. According to the central limit theorem, $z(t)$ has a normal distribution with mean 0 and variance t, and has independent and stationary increments. These are statistical properties described by equation (1.11).

1.2.7. Stationarity and ergodicity

These two terms have been frequently come across, relevant and important in financial and economic time series. Nonetheless, it is helpful here to provide simple definitions to link and distinguish them, and to clarify each of them. A stochastic process is said to be covariance stationary if:

(i) $E\{X(t)\} = \mu$ for all t;
(ii) $\mathrm{Var}\{X(t)\} < \infty$ for all t; and
(iii) $\mathrm{Cov}\{X(t), X(t+j)\} = \gamma_j$ for all t and j.

This is sometimes referred to as weekly stationary, or simply stationary. Such stationary processes have finite mean, variance and covariance that do not depend on the time t, and the covariance depends only on the interval j.

A strictly stationary process has met the above conditions (i) and (iii), and been extended to higher moments or orders. It states that the random vectors $\{X(t_1), X(t_2), \dots X(t_n)\}$ and $\{X(t_1+j), X(t_2+j), \dots X(t_n+j)\}$ have the same joint distribution. In other words, the joint distribution depends only on the interval j but not the time t. That is, the joint probability density $p\{x(t), x(t+\tau_1), \dots x(t+\tau_n)\}$, where $\tau_i = t_i - -t_{i-1}$, depends only on the intervals $\tau_1, \dots \tau_n$ but not t itself. A second-order stationary process is not exactly covariance stationary as it is not required to meet condition (ii). Therefore, a process can be strictly stationary while being not covariance stationary, and vice versa.

Ergodicity arises from the practical need to obtain ensemble moments' values from a single realisation or observation of the stochastic process. A covariance stationary process is ergodic for the first moment if its temporal average converges, with probability 1, to the ensemble average. Similarly, a covariance stationary process is ergodic for the second moment if its temporal covariance converges, with probability 1, to the ensemble covariance.

1.3. The behaviour of financial variables and beyond

A Wiener process has a mean value of zero and a unity variance. It is also a special type of random walk. The Wiener process can be generalised to describe a time

series where the mean value is a constant and can be different from zero, and the variance is a constant and can be different from unity. Most financial securities' prices fall in this category when the financial market is efficient in its weak form. An Ito process further relaxes these conditions so that both the deterministic and stochastic parts of the generalised Wiener process are state and time dependent. Important relationships between stochastic variables and, in particular, between a financial security's price and the price of its derivative, are established by Ito's lemma. Ito's lemma is central to the valuation and pricing of derivative securities, though it may shed light on issues beyond the derivative arena.

1.3.1. Generalised Wiener processes

A Wiener process described by equation (1.11) is a special and rather restricted random walk. It can be generalised so that the variance can differ from $1 \times t$ and there can be a drift. A stochastic process or variable x is a generalised Wiener process if:

$$\Delta x = a\Delta t + b\Delta z \tag{1.13}$$

where a is the drift rate, and b is the variance rate. Many financial time series can be subscribed to equation (1.13), especially in the context of so-called weak-form market efficiency, though equation (1.13) is a stronger claim to weak-form market efficiency than martingales.

1.3.2. Ito processes

If parameters a and b are functions of x and t, then equation (1.13) becomes the Ito process:

$$\Delta x = a(x,t)\,\Delta t + b(x,t)\,\Delta z \tag{1.14}$$

Function $a(x,t)$ can introduce the autoregressive component by having lagged Δx in it. Moving average effects can be introduced by $b(x,t)$ when it has non-zero constant values at times $t-i$ $(i = 1, 2, \ldots)$. Function $b(x,t)$ can generally introduce similar effects in the second moment, widely known as ARCH, GARCH, variations and stochastic volatility. Both $a(x,t)$ and $b(x,t)$ can bring in time varying coefficients in the first and second moments as well. Therefore, equation (1.14) can virtually represent all univariate time series found in finance and economics.

1.3.3. Ito's lemma

Ito's lemma is one of the most important tools for derivative pricing. It describes the behaviour of one stochastic variable as a function of another stochastic variable. The former could be the price of an option or the price of other derivatives, and the latter could be the price of shares.

Let us write equation (1.14) in the continuous time:

$$dx = a(x, t)\, dt + b(x, t)\, dz \tag{1.15}$$

Let y be a function of stochastic process x, Ito's lemma tells us that y is also an Ito process:

$$dy = \left(\frac{\partial y}{\partial x} a + \frac{\partial y}{\partial t} + \frac{1}{2} \frac{\partial^2 y}{\partial x^2} b^2 \right) dt + \frac{\partial y}{\partial x} b\, dz \tag{1.16}$$

It has a drift rate of:

$$\frac{\partial y}{\partial x} a + \frac{\partial y}{\partial t} + \frac{1}{2} \frac{\partial^2 y}{\partial x^2} b^2 \tag{1.17}$$

and a variance rate of:

$$\left(\frac{\partial y}{\partial x} \right)^2 b^2 \tag{1.18}$$

Equation (1.16) is derived by using the Taylor series expansion and ignoring higher orders of 0, details can be found in most mathematics texts at the undergraduate level.

Ito's lemma has a number of meaningful applications in finance and econometrics. Beyond derivative pricing, it reveals why and how two financial or economic time series are related to each other. For example, if two non-stationary (precisely, integrated of order 1) time series share the same stochastic component, the second term on the right-hand side of equations (1.15) and (1.16), then a linear combination of them is stationary. This phenomenon is called cointegration in the sense of Engle and Granger (1987) and Johansen (1988) in the time series econometrics literature. The interaction and link between them are most featured by the existence of an error correction mechanism. If two non-stationary time series are both the functions of an Ito's process, then they have a common stochastic component but may in addition have individual stochastic components as well. In this case, the two time series have a common trend in the sense of Stock and Watson (1988) but they are not necessarily cointegrated. This analysis can be extended to deal with stationary cases, e.g. common cycles in Engle and Issler (1995) and Vahid and Engle (1993).

1.3.4. *Geometric Wiener processes and financial variable behaviour in the short-term and long-run*

We can subscribe a financial variable, e.g. the share price, to a random walk process with normal distribution errors:

$$P_{t+1} = P_t + v_t, \qquad v_t \sim N(0, \sigma_P^2) \tag{1.19}$$

More generally, the price follows a random walk with a drift:

$$P_{t+1} = P_t + \phi + v_t, \qquad v_t \sim N(0, \sigma_P^2) \tag{1.20}$$

where ϕ is a constant indicating an increase (and less likely, a decrease) of the share price in every period. Nevertheless, a constant absolute increase or decrease in share prices is also not quite reasonable. A realistic representation is that the relative increase of the price is a constant:

$$\frac{P_{t+1} - P_t}{P_t} = \mu + \xi_t, \qquad \xi_t \sim N(0, \sigma^2) \tag{1.21}$$

So:

$$\Delta P_t = P_{t+1} - P_t = \mu P_t + P_t \xi_t = \mu P_t + \sigma P_t \varepsilon$$
$$\varepsilon \sim N(0, 1) \tag{1.22}$$

Notice $\Delta t = t + 1 - t = 1$ can be omitted in or added to the equations. Let Δt be a small interval of time (e.g. a fraction of 1), then equation (1.22) becomes:

$$\Delta P_t = \mu P_t \Delta t + \sigma P_t \varepsilon \sqrt{\Delta t} \tag{1.23}$$
$$= \mu P_t \Delta t + \sigma P_t \Delta z$$

Equation (1.23) is an Ito process in that its drift rate and variance rate are functions of the variable in concern and time. Applying Ito's lemma, we obtain the logarithm of the price as follows:

$$\Delta p_t = p_{t+1} - p_t = \left[\mu - \frac{\sigma^2}{2} \right] \Delta t + \Delta z \tag{1.24}$$

where $p_t = \ln(P_t)$ has a drift rate of $\mu' = \mu - (\sigma^2/2)$ and variance rate of σ^2. Equation (1.24) is just a generalised Wiener process instead of an Ito process in that its drift rate and variance rate are not the functions of P_t and t. This simplifies analysis and valuation empirically.

If we set $\sigma = 0$, the process is deterministic and solution is:

$$P_t = P_0(1 + \mu)^t \approx P_0 e^{\mu t} \tag{1.25}$$

and

$$p_t = p_0 + t \ln(1 + \mu) \approx p_0 + \mu t \tag{1.26}$$

The final result in equations (1.25) and (1.26) is obtained when μ is fairly small and is also the continuous time solution. From above analysis we can conclude that share prices grow exponentially while log share prices grow linearly.

When $\sigma \neq 0$, rates of return and prices deviate from above-derived values. Assuming there is only one shock (innovation) occurring in the kth period, $\varepsilon(k) = \sigma$, then:

$$P_t = P_0(1+\mu)(1+\mu)\cdots(1+\mu+\sigma)\cdots(1+\mu)(1+\mu)(1+\mu) \qquad (1.27)$$

for the price itself, and

$$p_t = p_0 + (t-1)\ln(1+\mu) + \ln(1+\mu+\sigma) \approx p_0 + \sigma + \mu t \qquad (1.28)$$

for the log price. After k, the price level increases by σ permanently (in every period after k). However, the rate of change or return is $\mu + \sigma$ in the kth period only; after k, the rate of return changes back to μ immediately after k.

The current rate of return or change does not affect future rates of return or change, so it is called a short-term variable. This applies to all similar financial and economic variables in the form of first difference. The current rate of return has an effect on future prices, either in original forms or logarithms, which are dubbed as long-run variables. Long-run variables often take their original form or are in logarithms, both being called variables in levels in econometric analysis. We have observed from above analysis that adopting variables in logarithms gives rise to linear relationships which simplify empirical analysis, so many level variables are usually in their logarithms.

In the above analysis of the share price, we reasonably assume that the change in the price is stationary and the price itself is integrated of order 1. Whereas under some other circumstances the financial variables in their level, not in their difference, may exhibit the property of a stationary process. Prominently, two of such variables are the interest rate and the unemployment rate. To accommodate this, a mean-reversion element is introduced in the process. Taking the interest rate for example, one of the models can have the following specification:

$$\Delta r_t = a(b - r_t)\Delta t + \sigma r_t \Delta z, \qquad a > 0, \quad b > 0 \qquad (1.29)$$

Equation (1.29) says that the interest rate decreases when its current value is greater than b and it increases when its current level is below b, where b is the mean value of the interest rate to which the interest rate reverts. A non-stationary process, such as that represented by equation (1.23), and a mean-reverse process, such as equation (1.29), differ in their statistical properties and behaviour. But more important are the differences in their economic roles and functions.

1.3.5. *Valuation of derivative securities*

In finance, Ito's lemma has been most significantly applied to the valuation of derivative securities, leading to so-called risk-neutral valuation principle. It can also be linked to various common factor analysis in economics and finance, notably cointegration, common trends and common cycles.

Let us write equation (1.23) in the continuous time for the convenience of mathematical derivative operations:

$$dP_t = \mu P_t \, dt + \sigma P_t \, dz \tag{1.30}$$

Let π_t be the price of a derivative security written on the share. According to Ito's lemma, we have:

$$d\pi_t = \left(\frac{\partial \pi_t}{\partial P_t} \mu P_t + \frac{\partial \pi_t}{\partial t} + \frac{1}{2} \frac{\partial^2 \pi_t}{\partial P_t^2} \sigma^2 P_t^2 \right) dt + \frac{\partial \pi_t}{\partial P_t} \sigma P_t \, dz \tag{1.31}$$

Now set up a portfolio which eliminates the stochastic term in equations (1.30) and (1.31):

$$\Omega_t = -\pi_t + \frac{\partial \pi_t}{\partial P_t} P_t \tag{1.32}$$

The change in Ω_t:

$$\begin{aligned} d\Omega_t &= -d\pi_t + \frac{\partial \pi_t}{\partial P_t} \, dP_t \\ &= \left(-\frac{\partial \pi_t}{\partial t} - \frac{1}{2} \frac{\partial^2 \pi_t}{\partial P_t^2} \sigma^2 P_t^2 \right) dt \end{aligned} \tag{1.33}$$

is deterministic involving no uncertainty. Therefore, Ω_t must grow at the risk-free interest rate:

$$d\Omega_t = r_f \Omega_t \, dt \tag{1.34}$$

where r_f is the risk-free interest rate. This shows the principle of risk neutral valuation of derivative securities. It should be emphasised that risk neutral valuation does not imply people are risk neutral in pricing derivative securities. In contrast, the general setting and background are that risk-averse investors make investment decisions in a risky financial world.

Substituting from equations (1.32) and (1.33), equation (1.34) becomes:

$$\left(\frac{\partial \pi_t}{\partial t} + \frac{1}{2} \frac{\partial^2 \pi_t}{\partial P_t^2} \sigma^2 P_t^2 \right) dt = r_f \left(\pi_t - \frac{\partial \pi_t}{\partial P_t} P_t \right) dt \tag{1.35}$$

$$\frac{\partial \pi_t}{\partial t} + \frac{\partial \pi_t}{\partial P_t} r_f P_t + \frac{1}{2} \frac{\partial^2 \pi_t}{\partial P_t^2} \sigma^2 P_t^2 = r_f \pi_t \tag{1.36}$$

Equation (1.36) establishes the price of a derivative security as the function of its underlying security and is a general form for all types of derivative securities. Combining with relevant conditions, such as the exercise price, time to maturity,

and the type of the derivative, a specific set of solutions can be obtained. It can be observed that solutions are much simpler for a forward/futures derivative, or any derivatives with their prices being a linear function of the underlying securities. It is because the third term on the left-hand side of equation (1.36) is zero for such derivatives.

Consider two derivative securities both written on the same underlying security such as a corporate share. Then, according to Ito's lemma, the two stochastic processes for these two derivatives subscribe to a common stochastic process generated by the process for the share price, and there must be some kind of fundamental relationship between them. Further, if two stochastic processes or financial time series are thought to be generated from or partly from a common source, then the two time series can be considered as being derived from or partly derived from a common underlying stochastic process, and can be fitted into the analytical framework of Ito's lemma as well. Many issues in multivariate time series analysis demonstrate this feature.

References

Engle, R.F. and Granger, C.W.J. (1987), Co-integration and error correction Representation, estimation, and testing, *Econometrica*, **55**, 251–267.

Engle, R.F. and Issler, J.V. (1995), Estimating common sectoral cycles, *Journal of Monetary Economics*, **35**, 83–113.

Jarrow, R.A. and Turnbull, S. (1999), *Derivative Securities* 2nd edn., South-Western College Publishing, Cincinnati, Ohio.

Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, **12**, 231–254.

Medhi, J. (1982), *Stochastic Processes*, Wiley Eastern, New Delhi.

Ross, S.M. (1996), *Stochastic Processes* 2nd edn., John Wiley, Chichester, England.

Stock, J.H. and Watson, M.W. (1988), Testing for common trends, *Journal of the American Statistical Association*, **83**, 1097–1107.

Vahid, F. and Engle, R.F. (1993), Common trends and common cycles, *Journal of Applied Econometrics*, **8**, 341–360.

2 Commonly applied statistical distributions and their relevance

This chapter reviews a few of statistical distributions most commonly assumed and applied in various subjects, including finance and financial economics. The first and foremost is the normal distribution. While introducing the normal distribution, we present it in an intuitive way, starting with few discrete states of events, through to more discrete states of events, and finally reaching the probability density function of normal distributions. The related concept of confidence intervals is also introduced, in conjunction with one of the financial market risk management measures, value at risk, so one can taste a flavour in finance and financial economics from the very beginning. The derivation, relevance and use of the χ^2-distribution, t-distribution and F-distribution are then presented and briefly discussed in sequence.

2.1. Normal distributions

The normal distribution is the most commonly assumed and applied statistical distribution. Many random variables representing various events are regarded normally distributed. Moreover, a few of other distributions are derived as some kinds of functions of normal distributions. A representation of the normal distribution, the probability density function of the normal distribution, is as follows:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{2.1}$$

Usually, $X \sim N(\mu, \sigma^2)$ is used to stand for a normal distribution with the mean being μ and the variance being σ^2, while $N(0, 1)$ is the standard normal distribution. Figure 2.1 illustrates normal distributions graphically, with two different variances, one small and one large.

Let us get some intuition behind the normal distribution. Suppose the economy is forecast to grow at 2 per cent per annum in the next year with a probability of 0.5, at 4 per cent with a probability of 0.25 and at 0 per cent with a probability of 0.25. The sum of all these probabilities of the states of events turning out to be true is surely 1, i.e. $0.5 + 0.25 + 0.25 = 1$. The top panel of Figure 2.2 illustrates the

Small standard deviation, σ Large standard deviation, σ

The whole area is one, whether σ is large or small.
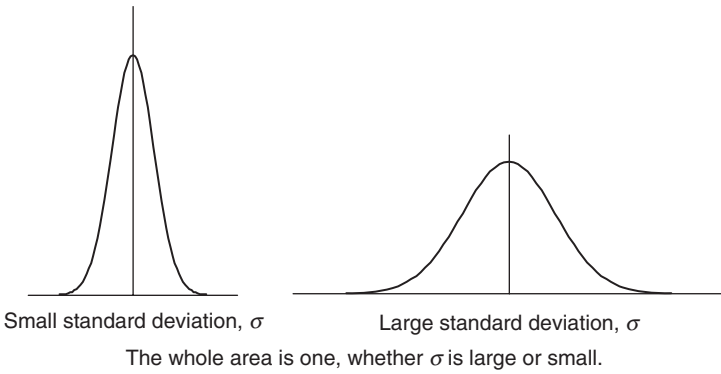
Figure 2.1 Normal distributions.
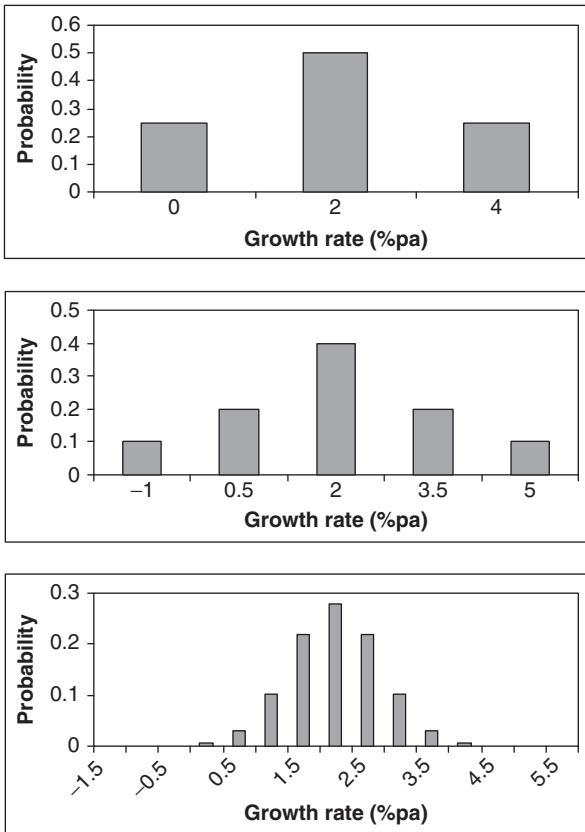
Figure 2.2 States of events: discrete but increase in numbers.

probable economic growth in the next year. With only three states of events, this is rather rough and sketchy both mathematically and graphically. So let us increase the states of events to five, the case illustrated by the middle panel of Figure 2.1. In that case, the economy is forecast to grow at 2 per cent with a probability of 0.4, at 3.5 per cent and 0.5 per cent with a probability of 0.2, and at 5 per cent and -1 per cent with a probability of 0.1. Similar to the first case, the sum of all the probabilities is $0.4 + 0.2 + 0.2 + 0.1 + 0.1 = 1$. This is more precise than the first case. The bottom panel of Figure 2.1 is a case with even more states of events, 15 states of events. The economy is forecast to grow at 2 per cent with a probability of 0.28, at 2.5 per cent and 1.5 per cent with a probability of 0.2181, at 3 per cent and 1 per cent with a probability of 0.1030, and so on. It looks quite like the normal distribution of Figure 2.1. Indeed, it is approaching normal distributions.

With a standard normal distribution, the probability that the random variable x takes values between a small interval $\Delta x = [x, x + \Delta x)$ is:

$$f(x) \cdot \Delta x = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \Delta x \tag{2.2}$$

This is the shaded area in Figure 2.3.

When $\Delta x \to 0$, the probability is:

$$f(x)\mathrm{d}x = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,\mathrm{d}x \tag{2.3}$$

While the probability density function of the standard normal distribution is defined as follows:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \tag{2.4}$$

Given any probability density function, the relationship between probability density function $f(x)$ and probability $P(x_1 \leq x < x_2)$ is:

$$P(x_1 < x \leq x_2) = \int_{x_1}^{x_2} f(x)\,\mathrm{d}x \tag{2.5}$$
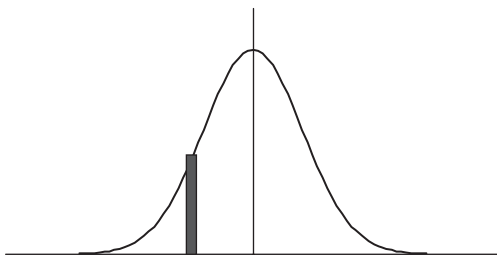
Figure 2.3 From discrete probabilities to continuous probability density function.

(a) Confidence interval

(b) Intuitive illustration of confidence intervals
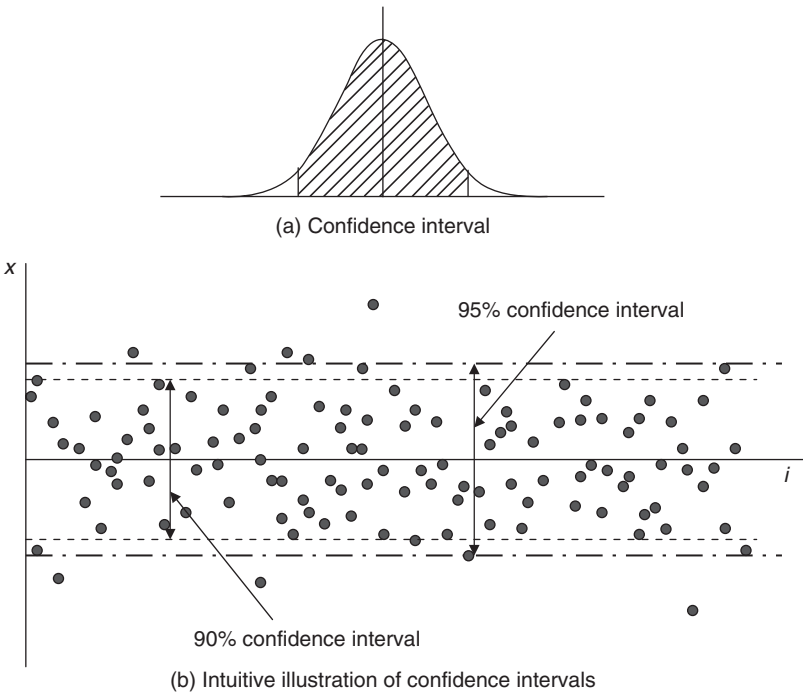
Figure 2.4 Illustrations of confidence intervals.

The confidence interval is to measure how sure an event or events would take place, given a criterion and with a probability. Figure 2.4 illustrates what are meant by confidence intervals. Part (a) of Figure 2.4 is based on a normal statistical distribution, with the shaded area being within a given confidence interval. A confidence interval is usually set to be sufficiently wide under the circumstances that the probability of the event taking place is large, typically 0.9 or 0.95, corresponding to the 90 per cent confidence interval and the 95 per cent confidence interval respectively, or even larger. Part (b) of Figure 2.4 is an intuitive illustration of confidence intervals without assuming any particular statistical distributions. Suppose it is an experiment to record rates of return on stocks for a sample of one hundred companies, with x, the vertical axis, being the rate of return and i, the horizontal axis, being the ith company. The 90 per cent confidence interval is an interval that about 90 companies' rates of return are contained by its two boundaries – above its lower boundary and below its upper boundary. The 95 per cent confidence interval is an interval that about 95 companies' rates of return are contained by its two boundaries – above its lower boundary and below its upper boundary. It is obvious that the 95 per cent confidence interval is wider than the 90 per cent confidence interval. Figure 2.5 portrays typical 90 per cent, 95 per cent and 98 per cent two-tailed confidence intervals, as well as 95 per cent, 97.5 per cent

and 99 per cent one-tailed confidence intervals, under normal distributions. These confidence intervals are also presented and explained numerically. For example, probability $(\mu - 1.96\sigma < x < \mu + 1.96\sigma) = 0.95$ in part (b) of Figure 2.5 indicates the meaning of the 95 per cent two-tailed confidence interval and its exact range under normal distributions. One may recall such frequently encountered numbers as 1.65, 1.96 and 2.33 in statistics, and inspecting part (d) to part (f) of Figure 2.5 may help one understand the meanings and relevance of statistical criteria here as well as in the latter part of this chapter.

Confidence intervals have various applications, e.g. they can be applied in significance tests such as whether two means are equal or from the same sample.

(a) 90% interval
two-tailed

90%

$\mu - 1.65\sigma$ μ $\mu + 1.65\sigma$ x

Probability $(\mu - 1.65\sigma < x < \mu + 1.65\sigma) = 0.9$
Or Probability $(x < \mu - 1.65\sigma) =$ Probability $(x > \mu + 1.65\sigma) = 0.05$

(b) 95% interval
two-tailed

95%

$\mu - 1.96\sigma$ μ $\mu + 1.96\sigma$ x

Probability $(\mu - 1.96\sigma < x < \mu + 1.96\sigma) = 0.95$
Or Probability $(x < \mu - 1.96\sigma) =$ Probability $(x > \mu + 1.96\sigma) = 0.025$

(c) 98% interval
two-tailed

98%

$\mu - 2.33\sigma$ μ $\mu + 2.33\sigma$ x

Probability $(\mu - 2.33\sigma < x < \mu + 2.33\sigma) = 0.98$
Or Probability $(x < \mu - 2.33\sigma) =$ Probability $(x > \mu + 2.33\sigma) = 0.01$
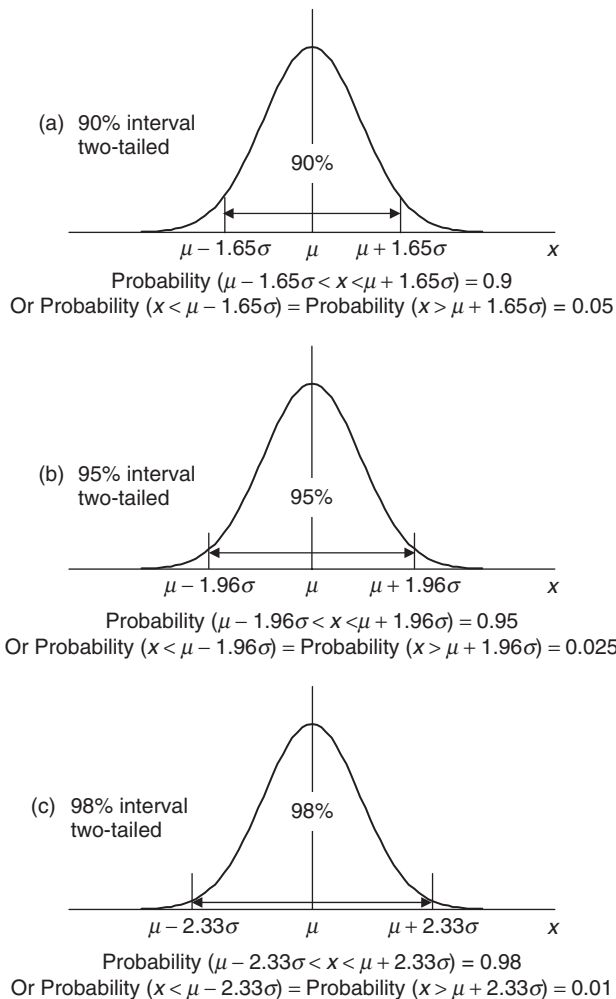
Figure 2.5 Two-tailed and one-tailed confidence intervals.

(d) 95% interval
one-tailed

95%

$\mu - 1.65\sigma$ μ x

Probability $(\mu - 1.65\ \sigma < x < \infty) = 0.95$
Or Probability $(x < \mu - 1.65\sigma) = 0.05$

(e) 97.5% interval
one-tailed

97.5%

$\mu - 1.96\sigma$ μ x

Probability $(\mu - 1.96\sigma < x < \infty) = 0.975$
Or Probability $(x < \mu - 1.96\sigma) = 0.025$

(f) 99% interval
one-tailed

99%

$\mu - 2.33\sigma$ μ $\mu + 2.33\sigma$ x

Probability $(\mu - 2.33\sigma < x < \infty) = 0.99$
Or Probability $(x < \mu - 2.33\sigma) = 0.01$
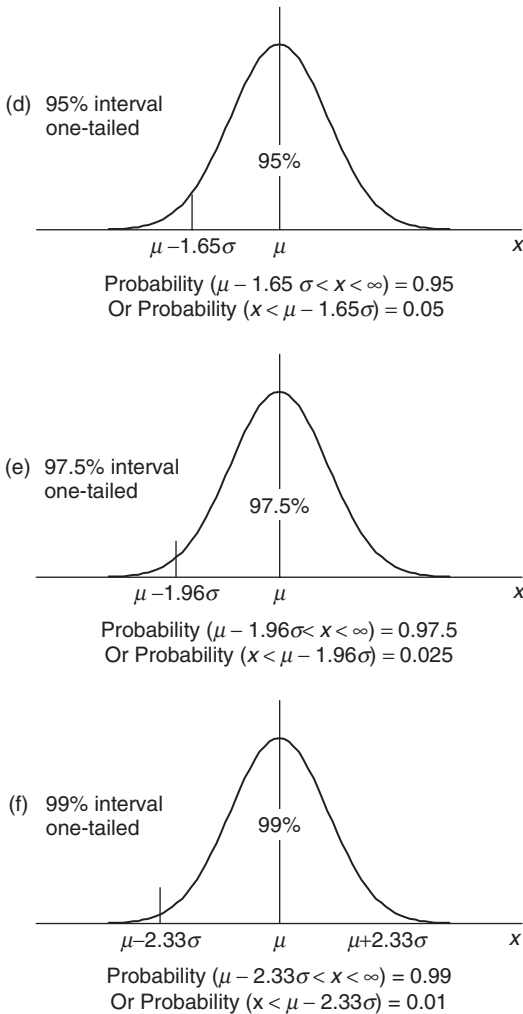
Figure 2.5 (continued)

Confidence intervals have much relevance in finance too. For example, Value at Risk (VaR) is an application of confidence intervals in market risk analysis, monitoring, control and management. VaR is a statistical and/or probability measure, against which the chance of a worse scenario happening is small (5 or 1 per cent). There is rationale in VaR. Prices soar and plummet, interest rates escalate and drop, exchange rates rise and fall. All of these take effect on the value of an asset, or a set of assets, a portfolio. How low the asset or portfolio value could become, or how much loss could be made, tomorrow, next week, next month, next year? Knowing the answers is very important to all financial

institutions and regulators. The following three examples show the application of VaR in relation to confidence intervals.

Example 2.1

Suppose the rate of return and the standard deviation of the rate of return on a traded stock are 10 per cent and 20 per cent per annum respectively; current market value of your investment in the stock is £10,000. What is the VaR over a one-year horizon, using 5 per cent as the criterion?

It is indeed an application of one-tailed 95 per cent confidence intervals, where $\mu = 10$ per cent, $\sigma = 20$ per cent, per annum. It can be easily worked out that $\mu - 1.65\sigma = 0.1 - 1.65 \times 0.2 = -0.23 = -23$ per cent. That is, there is a 5 per cent chance that the annual rate of return would be -23 per cent or lower. Therefore there is a 5 per cent chance that the asset value would be £10,000 $\times (1 - 23$ per cent$) = £7,700$ or lower in one year. £7,700 is the VaR over a one-year horizon. The result can be interpreted as follows. There is a 5 per cent chance that the value of your investment in the stock will be equal to or lower than £7,700 in one year; or there is a 95 per cent chance that that value will not be lower than £7,700 in one year. Another interpretation is that there is a 5 per cent chance that you will lose £10,000 − £7,700 = £2,300 or more in one year; or there is a 95 per cent chance you will not lose more than £2,300 in one year.

Example 2.2

Using the same information as in the previous example, what is the VaR over a one-day horizon with the 5 per cent criterion?

We adopt 250 working days per year, then $\mu = 0.1/250 = 0.0004$, $\sigma = 0.2/(250)^{0.5} = 0.012649$; $\mu - 1.65\sigma = 0.0004 - 1.65 \times 0.012649 = -0.02047 = -2.047$ per cent. This result indicates that there is a 5 per cent chance that the daily rate of return would be -2.047 per cent or lower. Therefore there is a 5 per cent chance that the asset value would be £10,000 $\times (1 - 2.047$ per cent$) = £9,795.29$ or lower in one day. £9,795.29 is the VaR over a one-day horizon. The result can be interpreted as follows. There is a 5 per cent chance that the value of your investment in the stock will be equal to or lower than £9,795.29 tomorrow; or there is a 95 per cent chance that that value will not be lower than £9,795.29 tomorrow. Another interpretation is that there is a 5 per cent chance that you will lose £10,000 − £9,795.29 = £204.71 or more tomorrow; or there is a 95 per cent chance you will not lose more than £204.71 tomorrow.

Example 2.3

This example demonstrates that the more volatile the rate of return, the lower is the VaR. If the standard deviation of the rate of return on the traded stock is 30 per cent per annum and all other assumptions are unchanged, what is the VaR over a one-day horizon, using 5 per cent as the criterion?

We still adopt 250 working days per year, then $\mu = 0.1/250 = 0.0004$, $\sigma = 0.3/(250)^{0.5} = 0.018974$; $\mu - 1.65\sigma = 0.0004 - 1.65 \times 0.018974 = -0.03091 = -3.091$ per cent. That is, there is a 5 per cent chance that the daily rate of return would be -3.091 per cent or lower. Therefore there is a 5 per cent chance that the asset value would be £10,000 × $(1 - 3.091$ per cent$) = £9,690.94$ or lower in one day. £9,690.94 is the VaR over a one-day horizon, which is lower compared with £9,795.29 when the $\sigma = 20$ per cent per annum in the previous case.

Before concluding this section, it is advisable to point out that the distributions of many economic and financial variables are lognormal instead of normal in their original forms. So let us present lognormal distributions and the transformation of variables briefly. The lognormal statistical distribution is described by the following formula:

$$f(x, m, s) = \frac{1}{x\sqrt{2\pi s^2}} e^{-(\ln x - m)^2/2s^2}$$

$$x > 0, \quad -\infty < m < \infty, \quad s > 0$$

(2.6)

This distribution is exhibited in Figure 2.6.

There are several reasons as to why a transformation is required. Firstly, many economic and financial variables grow exponentially, so their path is non-linear.
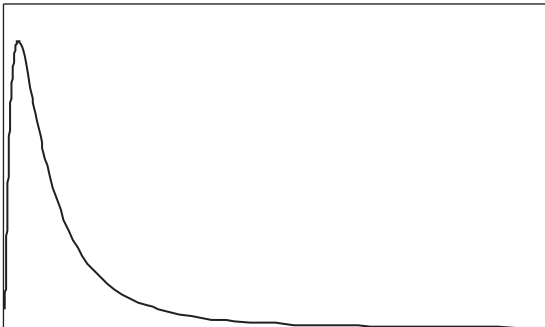
Figure 2.6 Lognormal distribution.

Transforming these variables with a logarithm operation achieves linearity. That is, the variables after the logarithm transformation grow linearly on a linear path. Secondly, the transformation through logarithm operations changes the variables in concern from absolute terms to relative terms, so comparison can be made cross-sections and over time. For example, the difference between the price in December and that in November is an absolute increase or change in price in one month; while the difference in the logarithm of the price in December and that in November is a relative increase or change in price in one month, i.e. a percentage increase or change. A monthly increase of £10 in one stock and a monthly increase of £12 in another stock cannot be directly compared. However, a 2 per cent monthly increase in one stock and a 1.5 per cent monthly increase in another clearly show the difference and superiority. Finally, the logarithm transformation may help achieve stationarity in time series data, though this statement may be controversial. Probably, one of the most convenient reasons for many economic and financial variables to follow lognormal distributions is the non-negative constraint. That is, these variables can only take values that are greater than or equal to zero. Due partly to this, values closer to zero are compressed and those far away from zero are stretched out. Lognormal distributions possess these features.

2.2. χ^2-distributions

The χ^2-distribution arises from the need in estimation of variance. It is associated with many test statistics, as they are about the variances under alternative specifications. It also leads to some other distributions, e.g. those involving both mean and variance.

The sum of the squares of m independent standard normal random variables obeys a χ^2-distribution with m degrees of freedom. Let Z_i ($i = 1, 2, \ldots m$) denote m independent $N(0, 1)$ random variables, then:

$$V = Z_1^2 + Z_2^2 + \cdots + Z_m^2 \sim \chi^2_{(m)} \tag{2.7}$$

The above random variable is said to obey a χ^2-distribution with m degrees of freedom. $\chi^2_{(m)}$ has a mean of m and a variance of $2m$, and is always non-negative. Figure 2.7 demonstrates several χ^2-distributions with different degrees of freedom. When m, the degree of freedom, becomes very large, the shape of the χ^2-distribution looks rather like a normal distribution.

The following case shows the need of χ^2-distributions in estimation of variance. Consider a sample:

$$Y_i = \beta + e_i, \qquad i = 1, \ldots T \tag{2.8}$$

where $Y_i \sim N(\beta, \sigma^2)$, $e_i \sim N(0, \sigma^2)$, and $\text{cov}(e_i, e_j) = 0$ for $j \neq i$.
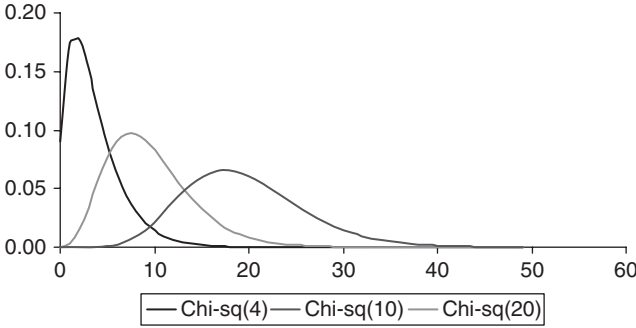
Figure 2.7 χ^2-distributions with different degrees of freedom.

An estimator of β is:

$$b = \frac{Y_1 + Y_2 + \cdots + Y_T}{T} = \frac{1}{T}\sum_{i=1}^{T} Y_i \tag{2.9}$$

and an estimator of σ^2 is:

$$\tilde{\sigma}^2 = \frac{e_1^2 + e_2^2 + \cdots + e_T^2}{T} = \frac{1}{T}\sum_{i=1}^{T} e_i^2 \tag{2.10}$$

Note that when e_i is unknown, an estimator of σ^2 is:

$$\tilde{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T} \hat{e}_i^2 \tag{2.11}$$

where the estimator of e_i is obtained through:

$$\hat{e}_i = Y_i - b \tag{2.12}$$

So, what kind of statistical distribution does $\tilde{\sigma}^2$ obey and what are the properties of the distribution? The answer is χ^2-distributions.

We can transform the errors in equation (2.8), resulting in:

$$\left(\frac{e_1}{\sigma}\right)^2 + \left(\frac{e_2}{\sigma}\right)^2 + \cdots + \left(\frac{e_T}{\sigma}\right)^2 = \sum_{i=1}^{T}\left(\frac{e_i}{\sigma}\right)^2 \sim \chi^2_{(T)} \tag{2.13}$$

and

$$\sum_{i=1}^{T}\left(\frac{\hat{e}_i}{\sigma}\right)^2 = \frac{T\tilde{\sigma}^2}{\sigma^2} \sim \chi^2_{(T-1)} \tag{2.14}$$

Due to loss of one degree of freedom in $\sum\limits_{i=1}^{T} \hat{e}_i = 0$, the degrees of freedom of the χ^2-distribution in equation (2.14) become $T - 1$.

The distribution of $\tilde{\sigma}^2$ is worked out as:

$$\tilde{\sigma}^2 \sim \frac{\sigma^2}{T} \chi^2_{(T-1)} \tag{2.15}$$

It is biased, since

$$E\left(\tilde{\sigma}^2\right) = \frac{\sigma^2}{T} E\left(\chi^2_{(T-1)}\right) = \frac{(T-1)}{T}\sigma^2 \neq \sigma^2$$

Correction is therefore required to obtain an unbiased estimator of σ^2. Notice:

$$E\left(\frac{T}{T-1}\tilde{\sigma}^2\right) = \frac{\sigma^2}{T-1} E\left(\chi^2_{(T-1)}\right) = \sigma^2$$

an unbiased estimator of σ^2 is:

$$\hat{\sigma}^2 = \frac{T}{T-1}\tilde{\sigma}^2 = \frac{1}{T-1}\sum_{i=1}^{T}\hat{e}_i^2 \sim \frac{\sigma^2}{T-1}\chi^2_{(T-1)} \tag{2.16}$$

as we can see that:

$$E\left(\frac{T}{T-1}\tilde{\sigma}^2\right) = \frac{\sigma^2}{T-1} E\left(\chi^2_{(T-1)}\right) = \sigma^2 \tag{2.17}$$

2.3. *t*-distributions

The *t*-distribution arises from the need in estimation of the accuracy of an estimate, or joint evaluation of the mean and variance of the estimate, or the acceptability of the estimate, when the variance is unknown. Many individual parameters, such as sample means and regression coefficients, obey *t*-distributions. It is a combination of two previously learned distributions, the normal distribution and the χ^2-distribution.

Let Z obey an $N(0, 1)$ distribution and χ^2_T follow a χ^2-distribution with T degrees of freedom, then:

$$t = \frac{Z}{\sqrt{\chi^2_{(T)}/T}} \sim t_{(T)} \tag{2.18}$$

The random variable of the above kind is said to obey a *t*-distribution with T degrees of freedom. Figure 2.8 shows *t*-distributions with different degrees
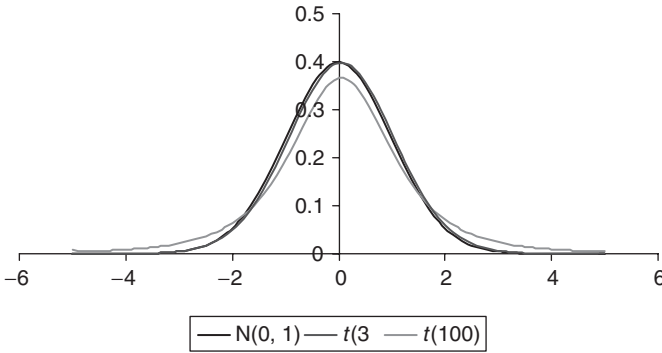
Figure 2.8 *t*-distributions with different degrees of freedom.

of freedom. When degrees of freedom become infinite, a *t*-distribution approaches a normal distribution.

Now let us consider the need in estimation of the accuracy of an estimate. Use the previous example of equation (2.8) and recall that an estimator of the sample mean is equation (2.9). The mean of the estimator represented by equation (2.9) is:

$$E(b) = E\left(\frac{1}{T}\sum_{i=1}^{T} Y_i\right) = \beta \tag{2.19}$$

Its variance is:

$$\mathrm{Var}(b) = \mathrm{Var}\left(\frac{1}{T}\sum_{i=1}^{T} Y_i\right) = \frac{\sigma^2}{T} \tag{2.20}$$

Therefore:

$$b \sim N\left(\beta, \frac{\sigma^2}{T}\right) \tag{2.21}$$

The above can be rearranged to:

$$\frac{b - \beta}{\sigma/\sqrt{T}} = Z \sim N(0, 1) \tag{2.22}$$

Equation (2.22) measures the 'meaningfulness' of *b*, or the closeness of *b* to *β*.

When σ^2 is unknown, we have to apply $\hat{\sigma}^2$, then:

$$b \sim N\left(\beta, \frac{\hat{\sigma}^2}{T}\right) \tag{2.23}$$

The measure of 'meaningfulness' becomes:

$$\frac{b - \beta}{\hat{\sigma}/\sqrt{T}} \tag{2.24}$$

Unlike the measure in equation (2.22), the above measure in equation (2.24) does not obey a normal distribution. Let us make some rearrangements:

$$\frac{b - \beta}{\hat{\sigma}/\sqrt{T}} = \frac{b - \beta}{\sigma/\sqrt{T}} \left/ \frac{\hat{\sigma}/\sqrt{T}}{\sigma/\sqrt{T}} \right. = \frac{b - \beta}{\sigma/\sqrt{T}} \left/ \sqrt{\frac{\hat{\sigma}^2}{\sigma^2}} \right. \tag{2.25}$$

Note in the above, the numerator is $N(0, 1)$, and the denominator is the square root of $\chi^2_{(T-1)}/(T-1)$, which leads to t-distributions with $(T-1)$ degrees of freedom.
From:

$$\frac{b - \beta}{\hat{\sigma}/\sqrt{T-1}} = t_{(T-1)} \tag{2.26}$$

the probability for b to be a reasonable estimator of β is given as:

$$P\left(-t_{\alpha/2,(T-1)}\hat{\sigma}/\sqrt{T-1} \le b - \beta \le t_{\alpha/2,(T-1)}\hat{\sigma}/\sqrt{T-1}\right) = 1 - \alpha \tag{2.27}$$

Equations (2.26) and (2.27) can be used to test the null hypothesis $H_0: b = \beta$, against the alternative $H_1: b \ne \beta$. This is called the t-test. Usually t-tests are one-tailed, with the one-sided test statistic:

$$P\left(-t_{\alpha,(T-1)}\hat{\sigma}/\sqrt{T-1} \le b - \beta \le \infty\right) = 1 - \alpha \tag{2.28}$$

The reasons for adopting one-tailed t-tests can be made clear from inspecting Figure 2.9. Part (a) of Figure 2.9 shows a small corner accounting for only 5 per cent of the whole area; the distance is 1.96 times of the standard deviation to the left of origin. If we shift the whole distribution to the right by this distance with the mean being the estimator of β, b, as shown with part (b) of Figure 2.9, then, a t-statistic of 1.96 means that there is only 5 per cent chance that $b \le 0$; or b is statistically different from zero at the 5 per cent significance level. The larger the t-statistic, the higher is the significance level. For example, with a t-statistic of 2.58, the chance for $b \le 0$ is even smaller, observed in part (c) of the Figure 2.9; and with a t-statistic of 3.75, the chance for $b \le 0$ is almost negligible, as demonstrated by part (d) of Figure 2.9.
Alternatively, the variable can be squared:

$$t^2_{(T-1)} = \frac{(b - \beta)^2}{\hat{\sigma}^2/T} = \frac{Z^2}{\chi^2_{(T-1)}/(T-1)} = \frac{\chi^2_{(1)}}{\chi^2_{(T-1)}/(T-1)} \tag{2.29}$$

It is the simplest of F-distributions to be discussed in the next section.
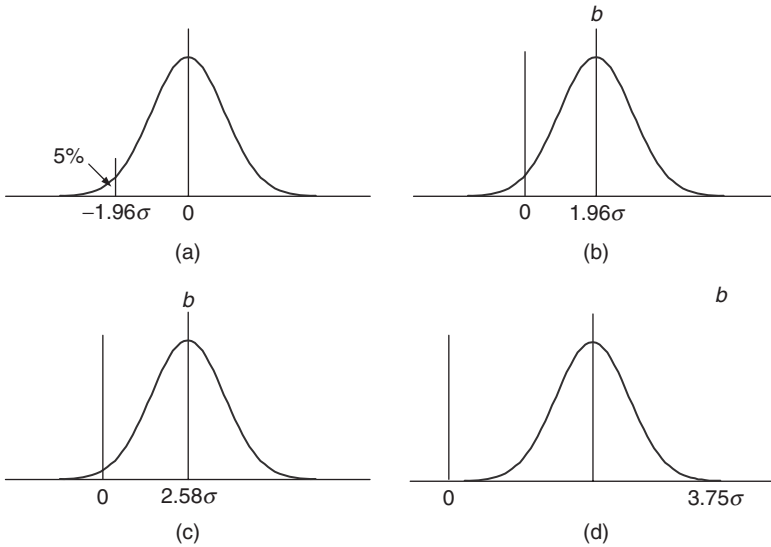
Figure 2.9 *t*-tests and the rationale.

2.4. *F*-distributions

The *F*-distribution is about the statistical distribution involving more than one parameter, an extension of the *t*-distribution. It is widely used in tests on the validity of one or more imposed restrictions. It is the ratio of two χ^2-distributions, i.e. the variance ratio of two specifications.

Let $\chi^2_{T_1}$ follow a χ^2-distribution with T_1 degrees of freedom and $\chi^2_{T_2}$ follow a χ^2-distribution with T_2 degrees of freedom, then:

$$F = \frac{\chi^2_{(T_1)}/(T_1)}{\chi^2_{(T_2)}/(T_2)} \sim F_{(T_1, T_2)} \tag{2.30}$$

The above random variable is said to follow an *F*-distribution with degrees of freedom T_1 and T_2.

The application and relevance of the *F*-distribution can be demonstrated with the following case. Some knowledge in linear regression and estimation methods such as the ordinary least squares may be required for a better understanding of this distribution and one may refer to the first section of the next chapter. Consider an extension of the previous example:

$$Y_i = \beta_1 + \beta_2 x_{2,i} + \cdots + \beta_K x_{K,i} + e_i, \quad i = 1, \ldots T \tag{2.31}$$

where $e_i \sim N(0, \sigma^2)$. We know the individual coefficients have a *t*-distribution:

$$\frac{b_j - \beta_j}{\hat{\text{Var}}(b)} \sim t_{(T-K)}, \quad j = 2, \ldots K \tag{2.32}$$

or:

$$t^2_{(T-K)} = \frac{(b_j - \beta_j)^2}{\text{Vâr}(b_j)} = \frac{\chi^2_{(1)}}{\chi^2_{(T-K)}/(T-K)}$$

$$= (b_j - \beta_j)\left[\text{Vâr}(b_j)\right]^{-1}(b_j - \beta_j) \tag{2.33}$$

Replacing the above single parameters by vectors and matrices, we have essentially obtained the joint distribution of the regression coefficients:

$$\frac{(\mathbf{b} - \boldsymbol{\beta})'\left[\mathbf{Côv}(\mathbf{b})\right]^{-1}(\mathbf{b} - \boldsymbol{\beta})}{K - 1} = \frac{\chi^2_{(K-1)}/(K-1)}{\chi^2_{(T-K)}/(T-K)} \tag{2.34}$$

where:

$$\mathbf{b} = \begin{bmatrix} b_2 & \cdots & b_K \end{bmatrix}'$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_2 & \cdots & \beta_K \end{bmatrix}'$$

$$\mathbf{Côv}(\mathbf{b}) = \begin{bmatrix} \text{Vâr}(b_2) & \cdots & \text{Côv}(b_1, b_K) \\ & \cdots & \\ \text{Côv}(b_K, b_1) & & \text{Vâr}(b_K) \end{bmatrix}$$

The above is an F-distribution with $(K-1)$ and $(T-K)$ degrees of freedom. The F-statistic is in fact the ratio of the explained variance and unexplained variance.

The t-statistic and the F-statistic are widely used to test restrictions, e.g. one, several or all regression coefficients are zero. Define SSE_U as the unrestricted sum of squared errors:

$$SSE_U = \hat{\mathbf{e}}'\hat{\mathbf{e}} = (\mathbf{y} - \mathbf{b}\mathbf{x})'(\mathbf{y} - \mathbf{b}\mathbf{x}) \tag{2.35}$$

SSE_R as the restricted sum of squared errors:

$$SSE_R = \hat{\mathbf{e}}^{*\prime}\hat{\mathbf{e}}^* = (\mathbf{y} - \mathbf{b}^*\mathbf{x})'(\mathbf{y} - \mathbf{b}^*\mathbf{x}) \tag{2.36}$$

With T observations, K regressors, and J restrictions, the F-test is:

$$\frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)} = F_{[J,(T-K)]} \tag{2.37}$$

If the restrictions are valid, then SSE_R will not be much larger than SSE_U, and the F-statistic will be small and insignificant. Otherwise, the F-statistic will be large and significant.

3 Overview of estimation methods

This chapter briefly reviews primary statistical estimation methods, firstly under the assumptions that the residuals from fitting the model obey independent identical distributions (iid) and the distributions may be further assumed to be normal, and then under more general conditions where iid of residuals is violated. For progression purposes, we firstly introduce the ordinary least squares method (OLS) and the maximum likelihood (ML) method under iid. Then, we relax the iid requirements and extend the OLS and the ML methods with certain modifications. Further relaxations of residual distribution requirements lead to general residual distributions typically observed in time series and cross-section modelling, which are addressed next. Finally, we present the estimation methods based on moment conditions, including the method of moments (MM) and the generalised method of moments (GMM), which are claimed to be more efficient, relaxed and easier in estimation processes and computation, and are getting momentum in applications in recent years.

3.1. Basic OLS procedures

Given a regression relation such as:

$$y_i = \beta_1 + \beta_2 x_i + e_i, \quad e_i \sim \text{iid}(0, \sigma^2), \quad i = 1, \dots N \tag{3.1}$$

the idea of the OLS is to minimise the sum of squared errors:

$$\sum_{i=1}^{N} e_i^2 = \sum_{i=1}^{N} (y_i - b_1 - b_2 x_i)^2 \tag{3.2}$$

The minimisation process is as follows. Taking derivatives of the above summation with respect to β_1 and β_2 respectively and setting them to zero:

$$-2\sum_{i=1}^{N}(y_i - b_1 - b_2 x_i) = 0$$

$$-2\sum_{i=1}^{N} x_i (y_i - b_1 - b_2 x_i) = 0 \tag{3.3}$$

Making rearrangements of the above equations leads to:

$$Nb_1 + b_2 \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i$$

$$b_1 \sum_{i=1}^{N} x_i + b_2 \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i$$

(3.4)

Then the OLS estimators of β_1 and β_2, b_1 and b_2, are solved as:

$$b_1 = \frac{N \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2}$$

(3.5)

$$b_2 = \frac{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i \right)^2}$$

(3.6)

The above analysis can be extended to multivariate cases, expressed in a compact form with vectors and matrices:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \qquad \mathbf{e} \sim \text{iid}\left(0, \sigma^2\right)$$

(3.7)

where $\mathbf{y}$ is an $(N \times 1)$ vector containing N observations of the dependent variable, X is an $(N \times K)$ matrix with K being the number of regressors including the intercept, i.e. $\mathbf{X}_i = \begin{bmatrix} 1 & x_{2,i} & \cdots & x_{k,i} \end{bmatrix}$, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of coefficients, and $\mathbf{e}$ is an $(N \times 1)$ vector of residuals.

The OLS procedure is to minimise:

$$(\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})$$

(3.8)

The minimisation leads to:

$$\mathbf{X}'(\mathbf{y} - \mathbf{Xb}) = 0$$

(3.9)

Then the estimated coefficients are obtained as:

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

(3.10)

The corresponding covariance matrix of the estimated coefficients is:

$$\mathbf{Cov(b)} = \hat{\sigma}^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

(3.11)

where:

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{e}}'\hat{\mathbf{e}}}{T - K} \tag{3.12}$$

and:

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{Xb} \tag{3.13}$$

3.2. Basic ML procedures

As long as a computable probability density function form can be figured out, the ML method and procedures can usually be applied. Nevertheless, the most commonly assumed function form is normal distributions or a function form that can be transformed into normal distributions. Given the same regression relation as in the previous section, but further assuming the error term is normally distributed, the model becomes:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$
$$\varepsilon_i \sim N(0, \sigma^2), \quad \mathrm{Cov}\left(\varepsilon_i, \varepsilon_j\right) = 0, \quad \text{for } j \neq i, \quad i = 1, \dots N \tag{3.14}$$

The ML method is to maximise the joint probability density function of normal distributions:

$$f(\varepsilon_1, \dots \varepsilon_N) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left[\varepsilon_1^2/2\sigma^2\right]} \dots \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left[\varepsilon_N^2/2\sigma^2\right]}$$
$$= \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left[(y_1 - b_1 - b_2 x_1)^2/2\sigma^2\right]} \dots \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\left[(y_N - b_1 - b_2 x_N)^2/2\sigma^2\right]} \tag{3.15}$$

Taking logarithms of the joint probability density function results in:

$$\mathrm{Ln}(f) = -N \cdot \mathrm{Ln}(\sigma) - \frac{N}{2}\mathrm{Ln}(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^{N} (y_i - b_1 - b_2 x_i)^2 \tag{3.16}$$

To maximise the above function is indeed to minimise the third term on the right-hand side, leading to the same results as in the OLS case. Extending the above analysis to the multivariate setting:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} \sim N(0, \sigma^2) \tag{3.17}$$

where all the descriptions, except that normality is imposed on $\boldsymbol{\varepsilon}$, are the same as before. The joint probability density function is:

$$f(\boldsymbol{\varepsilon}) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N e^{-\left[\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/2\sigma^2\right]} = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N e^{-\left[(\mathbf{y}-\mathbf{Xb})'(\mathbf{y}-\mathbf{Xb})/2\sigma^2\right]} \tag{3.18}$$

The logarithm of the above probability density function is:

$$\text{Ln}(f) = -N \cdot \text{Ln}(\sigma) - \frac{N}{2}\text{Ln}(2\pi) - \frac{(\mathbf{y} - \mathbf{Xb})'(\mathbf{y} - \mathbf{Xb})}{2\sigma^2} \tag{3.19}$$

Similar to the simple ML case represented by equations (3.14) to (3.16), to maximise the above function is indeed to minimise the third term on the right-hand side of the above equation, leading to the same results as in the OLS case. The corresponding covariance matrix of the estimated coefficients is also the same as in the OLS case. It can be concluded that, under the assumption of normal distributions of errors, the OLS and the ML yield the same estimation results for regression parameters.

3.3. Estimation when iid is violated

Violation of iid distributions generally takes place in the forms of heteroscedasticity, cross-correlation and serial correlation or autocorrelation. Heteroscedasticity means that the variances of errors are not equal between them, as against homoscedastic, identical distribution. In this section, we just use one simple example each to illustrate what heteroscedasticity and serial correlation are respectively and their consequences. We leave more general residual distributions to be dealt with in the next section of this chapter, where elementary time series and cross-section modelling will be briefly introduced and discussed, based on which and from which most models in the following chapters of this book are developed and evolved. For the simple case of heteroscedasticity, suppose in the following regression equation the variance is proportional to the size of the regressor:

$$y_i = \beta_1 + \beta_2 x_i + e_i$$
$$\text{Var}(e_i) = \sigma^2 x_i, \qquad \text{Cov}(e_i, e_j) = 0, \quad for\ j \neq i,\ i = 1, \dots N \tag{3.20}$$

It is assumed, though, $\text{Cov}(e_i, e_j) = 0$, for $j \neq i$ still holds. That is, there is no cross-correlation if i and j represent different units with data being collected at the same time or having no time implications; or there is no autocorrelation if i and j represent different time points. To deal with this heteroscedasticity, the OLS can be generalised to achieve iid distributions. One way to achieve iid is to divide both sides of the equation by $\sqrt{x_i}$, leading to a new regression relation:

$$y_i^* = \beta_1 x_{1,i}^* + \beta_2 x_{2,i}^* + e_i^* \tag{3.21}$$

where $y_i^* = y_i / \sqrt{x_i}$, $x_{1,i}^* = 1/\sqrt{x_i}$, $x_{2,i}^* = \sqrt{x_i}$, and $e_i^* = e_i / \sqrt{x_i}$.

Heteroscedasticity has been removed after the transformation, proven by the following:

$$\text{Var}(e_i^*) = \frac{\text{Var}(e_i)}{x_i} = \sigma^2 \tag{3.22}$$

We can now apply the OLS to the transformed regression equation. The above procedure is one type of the generalised least squares (GLS), or the weighted least squares (WLS). One of the constraints is that the value of the scaling regressor must be non-negative.

With heteroscedastic error terms and further assuming normality, the corresponding joint probability density function for equation (3.20) is:

$$
\begin{aligned}
f(\varepsilon_1, \dots \varepsilon_N) &= \frac{1}{\sigma\sqrt{2\pi x_1}}\, \mathrm{e}^{-\left[\varepsilon_1^2/2x_1\sigma^2\right]} \dots \frac{1}{\sigma\sqrt{2\pi x_N}}\, \mathrm{e}^{-\left[\varepsilon_N^2/2x_N\sigma^2\right]} \\
&= \frac{1}{\sigma\sqrt{2\pi x_1}}\, \mathrm{e}^{-\left[(y_1-b_1-b_2x_1)^2/2x_1\sigma^2\right]} \\
&\quad \dots \frac{1}{\sigma\sqrt{2\pi x_N}}\, \mathrm{e}^{-\left[(y_N-b_1-b_2x_N)^2/2x_N\sigma^2\right]}
\end{aligned}
\tag{3.23}
$$

We can maximise the above joint probability density function without transformation, though a transformation would correspond to that with the GLS more closely. Notice the original probability density function is:

$$
f(\varepsilon_i)\,\mathrm{d}\varepsilon_i = \frac{1}{\sigma\sqrt{2\pi x_i}}\, \mathrm{e}^{-\left[\varepsilon_i^2/2x_i\sigma^2\right]}\mathrm{d}\varepsilon_i
\tag{3.24}
$$

where the variances are heteroscedastic. Let $\varepsilon_i^* = \varepsilon_i/\sqrt{x_i}$, then $\mathrm{d}\varepsilon_i^* = \mathrm{d}\varepsilon_i/\sqrt{x_i}$, and the above function becomes a homoscedastic normal distribution:

$$
f(\varepsilon_i^*)\,\mathrm{d}\varepsilon_i^* = \frac{1}{\sigma\sqrt{2\pi}}\, \mathrm{e}^{-\left[\varepsilon_i^{*2}/2\sigma^2\right]}\mathrm{d}\varepsilon_i^*
\tag{3.25}
$$

One of the advantages of the ML method is that one can do estimation without transformation, as long as the probability density function can be written down explicitly.

Let us turn to autocorrelation now. If in the following regression:

$$
y_t = \beta_1 + \beta_2 x_t + e_t, \qquad t = 1, \dots T
\tag{3.26}
$$

$\mathrm{Cov}(e_t, e_{t-k}) = \rho_k\sigma^2 \neq 0$, with t representing time, then it is said that there is autocorrelation of order k in the residual. A simple case is the first-order autocorrelation where $\mathrm{Cov}(e_t, e_{t-1}) = \rho_1\sigma^2 \neq 0$, and $\mathrm{Cov}(e_t, e_{t-k}) = 0$ for $k \geq 2$:

$$
\begin{aligned}
y_t &= \beta_1 + \beta_2 x_t + e_t \\
e_t &= \rho_1 e_{t-1} + v_t, \qquad \mathrm{Cov}(v_t, v_\tau) = 0 \quad \text{for } \tau \neq t, \quad t = 1, \dots T
\end{aligned}
\tag{3.27}
$$

The following rearrangements can remove such autocorrelation. Lag equation (3.27) by one period and then multiply both sides by ρ_1:

$$
\rho_1 y_{t-1} = \rho_1 \beta_1 + \rho_1 \beta_2 x_{t-1} + \rho_1 e_{t-1}
\tag{3.28}
$$

Then, subtracting both sides of equation (3.28) from the corresponding sides of equation (3.27) leads to:

$$y_t - \rho_1 y_{t-1} = \beta_1 - \rho_1 \beta_1 + \beta_2 x_t - \rho_1 \beta_2 x_{t-1} + v_t \tag{3.29}$$

We can perform transformations such as $y_t^* = y_t - \rho_1 y_{t-1}$ and $x_t^* = x_t - \rho_1 x_{t-1}$ to remove autocorrelation.

A simpler practice is to include the lagged variables on the right-hand side so the residual becomes iid, as shown below:

$$y_t = \mu + \rho_1 y_{t-1} + \alpha_1 x_t + \alpha_2 x_{t-1} + v_t \tag{3.30}$$

Then the OLS or ML can be easily applied to obtain parameter estimates.

3.4. General residual distributions in time series and cross-section modelling

This section briefly presents elementary time series and cross-section modelling with two interrelated purposes. One is to introduce more general types of statistical distributions of residuals commonly found with statistical data sets, most, if not all of them, are either time series or cross-sectional data or a combination of time series and cross-sectional data. The other, interrelated with the first and already mentioned in the first, is to introduce elementary time series and cross-section modelling, because most models in the following chapters of this book are developed and evolved based on and from these elementary time series and cross-section models and specifications.

In short, this section deals with correlation, both in time series and in cross-sections. Let us address correlation in time series first. A common finding in time series regression is that the residuals are correlated with their own lagged values. Since this correlation is sequential in time, it is then called serial correlation. This serial correlation violates the standard assumption of regression theory that disturbances are not correlated with other disturbances. There are a number of problems associated with serial correlation. The OLS is no longer efficient among linear estimators. Since prior residuals help to predict current residuals, we can take advantage of this information to form a better prediction of the dependent variable. Furthermore, standard errors computed using the classic OLS formula are not correct, and are generally understated. Finally, if there are lagged dependent variables on the right-hand side, OLS estimators are biased and inconsistent.

Serial correlation takes place for various reasons and can take different forms. Firstly, effects of shocks do not realise immediately. Technical diffusion as a process typically possesses this feature. Inefficient financial markets, especially in the weak form, are linked to serial correlation in statistical terms. Liquidity constraints imply some corresponding action cannot be taken immediately and the delay brings about serial correlation. Recent development in bounded rationality suggests that human beings' expectations are not always rational; human beings

do not always behave rationally, not because they do not want to but because they have limitations. This is also a cause of serial correlation. Secondly, data collection and processing processes may generate serial correlation. For example, data collected at different time points at different constituent units may cause serial correlation. Averaging monthly data to form quarterly data, or in general, averaging higher frequency data to form lower frequency data introduces serial correlation. Contrary to intuitive belief, using more data does not automatically lead to the benefit of information utilisation. In many cases, it is more desirable to simply sample, say, every first month in a quarter to represent that quarter, than to use all the months in a quarter for that quarter.

Serial correlation can take the form of autoregression (AR), moving average (MA) and their combinations. In an AR process, the residual in the current period is correlated to the residual in the previous periods. For example, the following is an AR process of order 1, expressed as AR(1):

$$y_t = \mu + e_t$$
$$e_t = \rho_1 e_{t-1} + v_t, \qquad \text{Cov}(v_t, v_\tau) = 0 \quad \text{for } \tau \neq t, \quad t = 1, \dots T \tag{3.31}$$

where μ is the mean value of the process, e_t is a disturbance term, or called unconditional residual, v_t is innovation in the disturbance, also known as the one-period-ahead forecast error or the prediction error. v_t is the difference between the actual value of the dependent variable and a forecast made on the basis of the independent variables and the past forecast errors. In general, an AR process of order p, AR(p), takes the form of:

$$y_t = \mu + e_t$$
$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \cdots + \rho_p e_{t-p} + v_t \tag{3.32}$$
$$\text{Cov}(v_t, v_\tau) = 0 \quad \text{for } \tau \neq t, \quad t = 1, \dots T$$

The above equations can be expressed as:

$$y_t = \mu \left(1 - \rho_1 - \rho_2 - \cdots - \rho_p\right) + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + v_t \tag{3.33}$$

That is, we can express y_t as a function of its past values. In general, there can be exogenous explanatory variables or no exogenous explanatory variable on the right hand side of the above equations. The above representations are simple but also typical for the illustration of AR processes, though exogenous explanatory variables are usually involved when we refer to AR models.

The other commonly used time series model involves MA processes. In an MA process, the current innovation, or forecast error, as well as its lagged values, enters the estimation and forecast equation. A simple case of MA processes of order 1, expressed as MA(1), is:

$$e_t = v_t + \theta v_{t-1}$$
$$\text{Cov}(v_t, v_{t-1}) = 0, \quad t = 1, \dots T \tag{3.34}$$

In general, an MA process of order q, MA(q), takes the form of:

$$e_t = v_t + \theta_1 v_{t-1} + \cdots \theta_q v_{t-q}$$
$$\text{Cov}(v_t, v_\tau) = 0 \quad \text{for } \tau \neq t, \quad t = 1, \ldots T$$

(3.35)

The autoregressive and moving average specifications can be combined to form an ARMA specification:

$$y_t = \mu + e_t + \rho_1 e_{t-1} + \cdots + \rho_p e_{t-p} + v_t + \theta_1 v_{t-1} + \cdots \theta_q v_{t-q}$$
$$\text{Cov}(v_t, v_\tau) = 0 \quad \text{for } \tau \neq t, \quad t = 1, \ldots T$$

(3.36)

The above is an ARMA (p, q) process, autoregression of order p and moving average of order q.

The family of univariate time series models are not complete without a third element, orders of integration. Recall one of the requirements for regression analysis is that the data to be analysed must be stationary. A time series may or may not be stationary. Taking difference operations usually can make a non-stationary time series become stationary. When a time series requires d times of differencing to become stationary, it is said to be integrated of order d, designated $I(d)$. A stationary time series is $I(0)$. A time series involving all the three elements is called autoregressive integrated moving average (ARIMA) process, consisting of these three parts. The first part is the AR term. Each AR term corresponds to the use of a lagged value of the unconditional residual in estimation and forecasting. The second part is the integration order term. A first-order integrated component means that the model is designed for the first difference of the original series. A second-order component corresponds to applying the second difference, and so on. The third part is the MA term. An MA forecasting model uses lagged values of the forecast error to improve the current forecast. A first-order moving average term uses the most recent forecast error, a second-order term uses the forecast error from the two most recent periods, and so on. An ARIMA(p, d, q) model can be expressed by the following representation:

$$\Delta^d y_t = \mu + e_t + \rho_1 e_{t-1} + \cdots + \rho_p e_{t-p} + v_t + \theta_1 v_{t-1} + \cdots \theta_q v_{t-q}$$
$$\text{Cov}(v_t, v_\tau) = 0 \quad \text{for } \tau \neq t, \quad t = 1, \ldots T$$

(3.37)

where Δ^d is a difference operator for d times of differencing.

ARIMA models can be estimated by applying the ML method when innovations or forecast errors are assumed to follow normal distributions and other numerical methods involving, sometimes, a large number of iterations. It was computer time-consuming in the past but is no longer a concern nowadays. As stated earlier, a common finding in time series regressions is that the residuals are correlated with their own lagged values. This serial correlation violates the standard assumption of regression theory that disturbances are not correlated with

other disturbances. Consequently, before one can use an estimated equation for statistical inference, e.g. hypothesis tests and forecasting, one should generally examine the residuals for evidence of serial correlation. For first-order serial correlation, there is the Durbin–Watson (DW) statistic. If there is no serial correlation, the DW statistic will be around 2. The DW statistic will fall below 2 if there is positive serial correlation (in the worst case, it will be near zero). If there is negative correlation, the statistic will lie somewhere between 2 and 4. Since first-order serial correlation is just a special case of serial correlation and is encompassed by higher order serial correlation, testing for higher-order serial correlation is more important, general and relevant. As a result, there are a few test statistics for higher-order serial correlation. Unlike the DW statistic, most, if not all, test procedures for higher-order serial correlation provide significance levels of the test statistic, which makes them more rigorous. Commonly adopted higher-order serial correlation test statistics include the Breusch–Godfrey Lagrange multiplier (LM) test and the Ljung–Box Q-statistic. If there is no serial correlation in the residuals, the autocorrelations and partial autocorrelations at all lags should be nearly zero, and all Q-statistics should be insignificant with large p-values. A test statistic for nth-order serial correlation will be labelled as $Q(n)$ and LM(n). They form part of the diagnostic tests for the chosen specification.

In cross-sectional data, most units are correlated in a number of ways, which may or may not result in the residuals from fitting a certain model being correlated. Nevertheless, no correlation is just a special case, so let us deal with the universal situation of correlation. Unlike time series where time sequence is relevant and important, correlation in cross-sections does not possess a time horizon. Such correlation is referred to as cross correlation, in a similar way in which correlation in time series is referred to as serial correlation.

Let us relax the assumption in equation (3.7) that the residuals are independent, so the model becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$
$$E(\mathbf{e}) = \mathbf{0}, \quad \text{Cov}(\mathbf{e}'\mathbf{e}) = \boldsymbol{\Sigma} \tag{3.38}$$

That is, the covariance matrix $\boldsymbol{\Sigma}$ is in general not a diagonal matrix $\mathbf{I}$ that its off-diagonal elements are in general non-zero, i.e. $\text{Cov}(e_i, e_j) = \sigma_{ij} = \rho_{ij}\sigma^2 \neq 0$, for $i \neq j$. The covariance matrix is illustrated as follows:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma^2 & \rho_{12}\sigma^2 & .. & \rho_{1N}\sigma^2 \\ \rho_{21}\sigma^2 & \sigma^2 & .. & \rho_{2N}\sigma^2 \\ .. & .. & .. & .. \\ \rho_{N1}\sigma^2 & \rho_{N2}\sigma^2 & .. & \sigma^2 \end{bmatrix} \tag{3.39}$$

On many occasions, $\rho_{ij} = \rho$, for all $i \neq j$, for cross-sectional data, since there is no time sequence in cross-sections. Then the covariance matrix becomes:

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & .. & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & .. & \rho\sigma^2 \\ .. & .. & .. & .. \\ \rho\sigma^2 & \rho\sigma^2 & .. & \sigma^2 \end{bmatrix} \tag{3.40}$$

The GLS procedure is to minimise:

$$(\mathbf{y} - \mathbf{Xb})' \, \Sigma^{-1} \, (\mathbf{y} - \mathbf{Xb}) \tag{3.41}$$

The minimisation leads to:

$$\mathbf{X}' \Sigma^{-1} (\mathbf{y} - \mathbf{Xb}) = 0 \tag{3.42}$$

Then the estimated coefficients are obtained as:

$$\mathbf{b} = \left(\mathbf{X}' \Sigma^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{y} \tag{3.43}$$

It is noted that the covariance matrix of equation (3.40) is not readily available and has to be estimated first. The usual practice is to apply the OLS first for the purpose of calculating the residuals, from which the covariance matrix can be derived. The estimation of the covariance matrix in the later chapters when the GLS is involved follows the same procedure.

Further assuming normal distributions for the residuals, equation (3.38) becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma}) \tag{3.44}$$

Its joint probability density function is:

$$f(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}^{-1} \left(\frac{1}{\sqrt{2\pi}} \right)^N \mathrm{e}^{-\left[\left(\boldsymbol{\varepsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \right) / 2 \right]} = \boldsymbol{\Sigma}^{-1} \left(\frac{1}{\sqrt{2\pi}} \right)^N$$
$$\mathrm{e}^{-\left[(\mathbf{y} - \mathbf{Xb})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{Xb}) / 2 \right]} \tag{3.45}$$

Taking logarithms of the above probability density function leads to:

$$\mathrm{Ln}(f) = \frac{1}{2} \left[-N \times \mathrm{Ln}(2\pi) + \mathrm{Ln} \left(\left| \boldsymbol{\Sigma}^{-1} \right| \right) - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right] \tag{3.46}$$

The maximisation of the above function is equivalent to minimising equation (3.41) with the GLS procedure, which produces the same parameter estimates.

3.5. MM and GMM approaches

The moment based methods are statistical approaches to estimating parameters that makes use of sample moment conditions. It derives parameter estimates by equating sample moments to unobserved population moments with the assumed statistical distribution and then solving these equations. These estimation methods and, in particular, the GMM, have been proven powerful to apply in practice, where and when either function forms or residual distributions cannot be explicitly expressed. The moment conditions are specified by the following model:

$$E\left\{f\left(\mathbf{w}, \boldsymbol{\beta}_0\right)\right\} = 0 \tag{3.47}$$

where $\mathbf{w}$ is a vector of observed variables, including the dependent variable, independent variables and possibly instrument variables, and $\boldsymbol{\beta}_0$ is a vector of true parameters. The model is said to be identified when there is a unique solution to the model with regard to the parameters:

$$E\{f\left(\mathbf{w}, \boldsymbol{\beta}\right)\} = 0 \quad \text{if and only if} \quad \boldsymbol{\beta} = \boldsymbol{\beta}_0 \tag{3.48}$$

For example in the following simple regression:

$$y_i = \beta_1 + e_i, \qquad e_i \sim \text{iid}(0, \sigma^2), \quad i = 1, \dots N \tag{3.49}$$

the moment condition is:

$$E\{f\left(\mathbf{w}, \boldsymbol{\beta}\right)\} = E\left\{y_i - \beta_1\right\} = 0 \tag{3.50}$$

The sample's first moment is:

$$\frac{1}{N}\sum_{i=1}^{N} y_i \tag{3.51}$$

Equating the population moment with the sample moment, b_1, the MM estimator of β_1, is derived as:

$$b_1 = \frac{1}{N}\sum_{i=1}^{N} y_i \tag{3.52}$$

The sample's second moment is:

$$\frac{1}{N}\sum_{i=1}^{N} y_i^2 \tag{3.53}$$

Equating the population moments with the sample moments, $\hat{\sigma}^2$, the MM estimator of the variance σ^2, is derived as:

$$\hat{\sigma}^2 = E(y_i - \beta_1)^2 = E(y_i)^2 - E^2(\beta_1) = \frac{1}{N}\left[\sum_{i=1}^{N} y_i^2 - \left(\sum_{i=1}^{N} y_i\right)^2\right] \quad (3.54)$$

Now consider a one variable regression model:

$$y_i = \beta_1 + \beta_2 x_i + e_i, \qquad e_i \sim \text{iid}(0, \sigma^2), \quad i = 1, \dots N \quad (3.55)$$

The moment conditions are:

$$E\{f(\mathbf{w}, \boldsymbol{\beta})\} = E\left\{\begin{bmatrix} 1 \\ x_i \end{bmatrix} e_i\right\} = E\left\{\begin{bmatrix} 1 \\ x_i \end{bmatrix} (y_i - \beta_1 - \beta_2 x_i)\right\} = 0 \quad (3.56)$$

where $\mathbf{w}$ includes y_i and x_i. That is, there are two moment conditions: $E\{(y_i - \beta_1 - \beta_2 x_i)\} = 0$ and $E\{x_i(y_i - \beta_1 - \beta_2 x_i)\} = 0$, and the parameter estimates can be derived just like the OLS procedure:

$$b_1 = \frac{N \sum_{i=1}^{N} x_i y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} y_i}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} \quad (3.57)$$

$$b_2 = \frac{\sum_{i=1}^{N} x_i^2 \sum_{i=1}^{N} y_i - \sum_{i=1}^{N} x_i \sum_{i=1}^{N} x_i y_i}{N \sum_{i=1}^{N} x_i^2 - \left(\sum_{i=1}^{N} x_i\right)^2} \quad (3.58)$$

The above analysis can be extended to multivariate cases, expressed in a compact form involving vectors and matrices:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \quad (3.59)$$

where $\mathbf{y}$ is an $(N \times 1)$ vector containing N observations of the dependent variable, $\mathbf{X}$ is an $(N \times K)$ matrix with K being the number of regressors including the intercept, i.e. $\mathbf{X}_i = \begin{bmatrix} 1 & x_{2,i} & \cdots & x_{k,i} \end{bmatrix}$, $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of coefficients, and $\mathbf{e}$ is an $(N \times 1)$ vector of residuals.

The K moment conditions are:

$$E\{\mathbf{X}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\} = 0 \quad (3.60)$$

which is indeed the results from minimising:

$$E\left\{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \tag{3.61}$$

and the estimated parameters are derived just like the OLS procedure:

$$\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \tag{3.62}$$

In general, the moment conditions can be written as:

$$E\left\{\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} = 0 \tag{3.63}$$

where $\mathbf{Z}$ is a vector of regressors that contains some of $\mathbf{X}$ and may contain instrument variables, and $\boldsymbol{\beta}$ is a vector of coefficients. When there are J moment conditions and J is equal to K, the above model is just-identified. If J is greater than K, the model is over-identified. The model is under-identified if J is smaller than K. Under-identification happens when one or more variables in vector $\mathbf{X}$ are endogenous and consequently:

$$E\left\{\mathbf{X}'_{\text{end}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \neq 0 \tag{3.64}$$

where $\mathbf{X}_{\text{end}}$ is a sub-set of $\mathbf{X}$ and are endogenous. Instruments are required when a sub-set of $\mathbf{X}$ are endogenous or correlated with the residual. Over-identification emerges when one or more instrument variables are employed, so that the number of moment conditions is greater than the number of variables in $\mathbf{X}$. This gives rise to the GMM. Let us define:

$$g(\boldsymbol{\beta}) = E\left\{\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right\} \tag{3.65}$$

There are two features of the GMM. Instead of choosing K moment conditions from the J moment conditions, the GMM uses all J moment conditions; then, instead of minimising the quadratic form of the moment conditions, the GMM minimises weighted quadratic moment conditions. The procedure for deriving GMM estimators are usually as follows. First formulate a quadratic form of the distance, or the weighted quadratic moment conditions:

$$q(\boldsymbol{\beta}) = g'(\boldsymbol{\beta})\mathbf{W}_g g(\boldsymbol{\beta}) \tag{3.66}$$

Then GMM estimators of $\boldsymbol{\beta}$, $\mathbf{b}$, are derived as:

$$\mathbf{b} = \arg\min_{\boldsymbol{\beta}} q(\boldsymbol{\beta}) = \arg\min_{\boldsymbol{\beta}} g'(\boldsymbol{\beta})\mathbf{W}_g g(\boldsymbol{\beta}) \tag{3.67}$$

where arg min is the global minimisation operation. The choice of weight matrices is important in implementing the GMM.

To actually implement the GMM, the two-step efficient GMM and iterated efficient GMM procedures are usually applied to derive GMM estimators of parameters. The two-step efficient GMM procedure works as follows. In the first step, an initial weight matrix that can be any arbitrary positive definite and symmetric matrix is chosen for the global minimisation operation. This initial weight matrix is usually an identity matrix. Therefore, the global minimisation process in the first step is amounted to:

$$\mathbf{b}^{[1]} = \arg \min_{\beta} q(\beta) = \arg \min_{\beta} g'(\beta) g(\beta) \tag{3.68}$$

The second step is to calculate a new weight matrix, and then derive GMM estimators using the new weight matrix. The second step weight matrix is estimated by:

$$\mathbf{W}_g^{[2]} = \mathbf{S}^{-1} \tag{3.69}$$

where

$$\mathbf{S} = \mathbf{S}_0 + \sum_{m=1}^{l} w_m \left(\mathbf{S}_m + \mathbf{S}'_m \right) \tag{3.70}$$

with w_m being the weight and

$$\mathbf{S}_m_{\substack{m=j-i}} = E \left\{ g_i(\theta) g'_j(\theta) \right\} \tag{3.71}$$

i and j in the above equation represent different time periods for time series data, and they represent different units for cross-sectional data. That is, serial correlation or cross-correlation is taken into account in the weight matrix. The second step estimators are derived through the following global minimisation:

$$\mathbf{b}^{[2]} = \arg \min_{\beta} q(\beta) = \arg \min_{\beta} g'(\beta) \mathbf{W}_g^{[2]} g(\beta)$$
$$= \arg \min_{\beta} g'(\beta) \mathbf{S}^{-1} g(\beta) \tag{3.72}$$

With the iterated efficient GMM procedure, the above two-step is repeated or iterated until convergence, i.e. when there is no significant difference in derived estimators from one iteration to the next.

GMM estimators possess the following asymptotic properties:

$$\sqrt{N} (\mathbf{b}_{\mathrm{GMM}} - \beta) \to N(\mathbf{0}, \mathbf{V}) \tag{3.73}$$

The asymptotic covariance matrix in the above distribution is:

$$\mathbf{V} = \left(\Omega' \mathbf{W} \Omega \right)^{-1} \left(\Omega' \mathbf{W} \mathbf{S} \mathbf{W} \Omega \right) \left(\Omega' \mathbf{W} \Omega \right)^{-1} \tag{3.74}$$

where:

$$\boldsymbol{\Omega} = E\left[\frac{\partial g\,(\mathbf{X}, \mathbf{Z}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right] = E\left[\mathbf{Z}'\mathbf{X}\right] \qquad (3.75)$$

As shown from equation (3.72), efficient GMM estimators are derived from setting the weight matrix to $\mathbf{S}^{-1}$, then the asymptotic covariance matrix for efficient GMM estimators are reduced to:

$$\mathbf{V} = \left(\boldsymbol{\Omega}'\mathbf{S}^{-1}\boldsymbol{\Omega}\right)^{-1} \qquad (3.76)$$

4 Unit roots, cointegration and other comovements in time series

The distinction between long-run and short-term characteristics in time series has attracted much attention in the last two decades. Long-run characteristics in economic and financial data are usually associated with non-stationarity in time series and called trends. Whereas short-term fluctuations are stationary time series and called cycles. Economic and financial time series can be viewed as combinations of these components of trends and cycles. Typically, a shock to a stationary time series would have an effect which would gradually disappear, leaving no permanent impact on the time series in the distant future. Whereas a shock to a non-stationary time series would permanently change the path of the time series; or permanently move the activity to a different level, either higher or a lower level.

Moreover, the existence of common factors among two or more time series may have such effect that the combination of these times series demonstrates no features which individual time series possess. For example, there could be a common trend shared by two time series. If there is no further trend which exists in only one time series, then it is said that these two time series are cointegrated. This kind of common factor analysis can be extended and applied to stationary time series as well, leading to the idea of common cycles.

This chapter first examines the properties of individual time series with regard to stationarity and tests for unit roots. Then, cointegration and its testing procedures are discussed. Finally, common cycles and common trends are analysed to further scrutinise comovements amongst variables.

4.1. Unit roots and testing for unit roots

Chapter 1 has provided a definition for stationarity. In the terminology of time series analysis, if a time series is stationary it is said to be integrated of order zero, or $I(0)$ for short. If a time series needs the difference operation once to achieve stationarity, it is an $I(1)$ series; and a time series is $I(n)$ if it is to be differenced for n times to achieve stationarity. An $I(0)$ time series has no roots on or inside the unit circle but an $I(1)$ or higher-order integrated time series contains roots on or inside the unit circle. So, examining stationarity is equivalent to testing for the existence of unit roots in the time series.

A pure random walk, with or without a drift, is the simplest non-stationary time series:

$$y_t = \mu + y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{4.1}$$

where μ is a constant or drift, which can be zero, in the random walk. It is non-stationary as $\mathrm{Var}(y_t) = t\sigma_\varepsilon^2 \to \infty$ as $t \to \infty$. It does not have a definite mean either. The difference of a pure random walk is the Gaussian white noise, or the white noise for short:

$$\Delta y_t = \mu + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{4.2}$$

The variance of Δy_t is σ_ε^2 and the mean is μ.

The presence of a unit root can be illustrated as follows, using a first-order autoregressive process:

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{4.3}$$

Equation (4.3) can be recursively extended, yielding:

$$y_t = \mu + \rho y_{t-1} + \varepsilon_t$$
$$= \mu + \rho\mu + \rho^2 y_{t-2} + \rho\varepsilon_{t-1} + \varepsilon_t$$
$$\cdots$$
$$= \left(1 + \rho + \cdots + \rho^{n-1}\right)\mu + \rho^n y_{t-n} + \left(1 + \rho L + \cdots + \rho^{n-1}L^{n-1}\right)\varepsilon_t \tag{4.4}$$

where L is the lag operator. The variance of y_t can be easily worked out:

$$\mathrm{Var}(y_t) = \frac{1 - \rho^n}{1 - \rho}\sigma_\varepsilon^2 \tag{4.5}$$

It is clear that there is no finite variance for y_t if $\rho \geq 1$. The variance is $\sigma_\varepsilon^2/(1 - \rho)$ when $\rho < 1$.

Alternatively, equation (4.3) can be expressed as:

$$y_t = \frac{\mu + \varepsilon_t}{(1 - \rho L)} = \frac{\mu + \varepsilon_t}{\rho((1/\rho) - L)} \tag{4.6}$$

which has a root $r = 1/\rho$.[1] Compare equation (4.5) with equation (4.6), we can see that when y_t is non-stationary, it has a root on or inside the unit circle, i.e. $r \geq 1$; while a stationary y_t has a root outside the unit circle, i.e. $r < 1$. It is usually said that there exists a unit root under the circumstances where $r \geq 1$. Therefore, testing for stationarity is equivalent to examining whether there is a unit root in the time series. Having gained the above idea, commonly used unit root test procedures are introduced and discussed in the following.

4.1.1. Dickey and Fuller

The basic Dickey–Fuller (DF) test (Dickey and Fuller, 1979, 1981) is to examine whether $\rho < 1$ in equation (4.3), which can, after subtracting y_{t-1} from both sides, be written as:

$$\begin{aligned}
\Delta y_t &= \mu + (\rho - 1)y_{t-1} + \varepsilon_t \\
&= \mu + \theta y_{t-1} + \varepsilon_t
\end{aligned} \tag{4.7}$$

The null hypothesis is that there is a unit root in y_t, or H_0: $\theta = 0$, against the alternative H_1: $\theta < 0$, or there is no unit root in y_t. The DF test procedure emerged since under the null hypothesis the conventional t-distribution does not apply. So whether $\theta < 0$ or not cannot be confirmed by the conventional t-statistic for the θ estimate. Indeed, what the Dickey–Fuller procedure gives us is a set of critical values developed to deal with the non-standard distribution issue, which are derived through simulation. Then, the interpretation of the test result is no more than that of a simple conventional regression.

Equations (4.3) and (4.7) are the simplest case where the residual is white noise. In general, there is serial correlation in the residual and Δy_t can be represented as an autoregressive process:

$$\Delta y_t = \mu + \theta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \varepsilon_t \tag{4.8}$$

Corresponding to equation (4.8), Dickey and Fuller's procedure becomes the augmented Dickey–Fuller test, or the ADF test for short. We can also include a deterministic trend in equation (4.8). Altogether there are four test specifications with regard to the combinations of an intercept and a deterministic trend.

4.1.2. Phillips and Perron

Phillips and Perron's (1988) approach is one in the frequency domain, termed as the PP test. It takes the Fourier transform of the time series Δy_t such as in equation (4.8), then analyses its component at the zero frequency. The t-statistic of the PP test is calculated as:

$$t = \sqrt{\frac{r_0}{h_0}} t_\theta - \frac{(h_0 - r_0)}{2h_0\sigma} \sigma_\theta \tag{4.9}$$

where

$$h_0 = r_0 + 2\sum_{\tau=1}^{M} \left(1 - \frac{j}{T}\right) r_j$$

is the spectrum of Δy_t at the zero frequency,[2] r_j is the autocorrelation function at lag j, t_θ is the t-statistic of θ, σ_θ is the standard error of θ, and σ is the standard error

of the test regression. In fact, h_0 is the variance of the M-period differenced series, $y_t - y_{t-M}$; while r_0 is the variance of the one-period difference, $\Delta y_t = y_t - y_{t-1}$.

Although it is not the purpose of the book to describe technical details of testing procedures, it is helpful to present intuitive ideas behind them. We inspect two extreme cases, one the time series is a pure white noise and the other a pure random walk. In the former, $r_j = 0$, $j \neq 0$ and $r_0 = h_0$, so $t = t_\theta$ and the conventional t-distribution applies. In the latter, $h_0 = M \times r_0$. If we look at the first term on the right-hand side of equation (4.9), t is adjusted by a factor of $\sqrt{1/M}$; and it is further reduced by value of the second term $\approx \sigma_\theta / 2\sigma$. So, the PP test gradually reduces the significance of the θ estimate as ρ moves from zero towards unity (or as θ moves from -1 to 0) to correct for the effect of non-conventional t-distributions, which becomes increasingly severe as ρ approaches unity.

4.1.3. Kwiatkowski, Phillips, Schmidt and Shin

Recently a procedure proposed by Kwiatkowski, Phillips, Schmidt and Shin (1992), known as the KPSS test named after these authors, has become a popular alternative to the ADF test. As the title of their paper, 'Testing the null hypothesis of stationarity against the alternative of a unit root', suggests, the test tends to accept stationarity, which is the null hypothesis, in a time series. Whereas in the ADF test the null hypothesis is the existence of a unit root, stationarity is more likely to be rejected. Many empirical studies have employed the KPSS procedure to confirm stationarity in such economic and financial time series as the unemployment rate and the interest rate, which, arguably, must be stationary for economic theories, policies and practice to make sense. Others, such as tests for purchasing power parity (PPP), are less restrictive by the theory. Confirmation or rejection of PPP is both acceptable in empirical research using a particular set of time series data, though different test results give rise to rather different policy implications. It is understandable that, relative to the ADF test, the KPSS test is less likely to reject PPP.

4.1.4. Panel unit root tests

Often in an empirical study, there is more than one time series to be examined. These time series are the same kind of data, such as the real exchange rate, current account balance or dividend payout, but they are for a group of economies or companies. These time series probably have the same length with the same start date and end date, or can be adapted without losing general properties. Under such circumstances, a test on pooled cross-section time series data, or panel data, can be carried out. Panel unit root tests provide an overall aggregate statistic to examine whether there exists a unit root in the pooled cross-section time series data and judge the time series property of the data accordingly. This, on the one hand, can avoid obtaining contradictory results in individual time series to which no satisfactory explanations can be offered. On the other hand, good asymptotic properties can be reached with relatively small samples in individual time series,

which are sometimes too small to be effectively estimated. In the procedure developed by Levin and Lin (1992, 1993), when the disturbances are i.i.d., the unit root t-statistic converges to the normal distribution; when fixed effects or serial correlation is specified for the disturbances, a straightforward transformation of the t-statistic converges to the normal distribution too. Therefore, their unit root t-statistic converges to the normal distribution under various assumptions about disturbances. Due to the presence of a unit root, the convergence is achieved more quickly as the number of time periods grows than as the number of individuals grows. It is claimed that the panel framework provides remarkable improvements in statistical power compared to performing a separate unit root for each individual time series. Monte Carlo simulations indicate that good results can be achieved in relatively small samples with 10 individual time series and 25 observations in each time series. Im *et al.* (1995) develop a $\bar{t}$ (t bar) statistic based on the average of the ADF t-statistics for panel data. It is shown that under certain conditions the $\bar{t}$-statistic has a standard normal distribution for a finite number of individual time series observations, as long as the number of cross-sections is sufficiently large. Commenting on and summarising Levin and Lin (1992, 1993) and Im *et al.* (1995) procedures, Maddala and Wu (1999) argue that the Levin and Lin test is too restrictive to be of interest in practice. While the test of Im *et al.* relaxes Levin and Lin's assumptions, it presents test results which merely summarise the evidence from a number of independent tests of the sample hypothesis. They subsequently suggest the Fisher test as a panel data unit root test and claim that the Fisher test with bootstrap-based critical values is the preferred choice.

4.2. Cointegration

Cointegration is one of the most important developments in time series econometrics in the last quarter century. A group of non-stationary $I(1)$ time series is said to have cointegration relationships if a certain linear combination of these time series is stationary. There are two major approaches to testing for cointegration, the Engle–Granger two-step method (Engle and Granger, 1987) and the Johansen procedure (Johansen, 1988, 1991; Johansen and Juselius, 1990). In addition, procedures for panel cointegration (Kao and Chiang, 1998; Moon and Phillips, 1999; Pedroni, 1999) have been recently developed, in the same spirit of panel unit roots and to address similar issues found in unit root tests. Since most panel cointegration tests employ the same estimation methods of or make minor adjustments in relation to the asymptotic theory of non-stationary panel data, they are not to be discussed in this chapter. The Engle–Granger method involves firstly running regression of one variable on another, and secondly checking whether the regression residual from the first step is stationary using, say, an ADF test. In this sense, the Engle–Granger method is largely the unit root test and will not be deliberated either. This chapter only presents the Johansen procedure which is to test restrictions imposed by cointegration on a VAR model:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots \mathbf{A}_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t \tag{4.10}$$

where $\mathbf{y}_t$ is a k-dimension vector of variables which are assumed to be $I(1)$ series (but can also be $I(0)$), $\mathbf{A}_i, i = 1, \ldots p$ is the coefficient matrix, and $\boldsymbol{\varepsilon}_t$ is a k-dimension vector of residuals. Subtracting $\mathbf{y}_{t-1}$ from both sides of equation (4.10) yields:

$$\Delta\mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Pi}\mathbf{y}_{t-1} + \boldsymbol{\Gamma}_1\Delta\mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta\mathbf{y}_{t-p+1} + \boldsymbol{\varepsilon}_t \qquad (4.11)$$

where:

$$\boldsymbol{\Pi} = \sum_{i=1}^{p}\mathbf{A}_i - \mathbf{I}$$

and:

$$\boldsymbol{\Gamma}_i = -\sum_{j=i+1}^{p}\mathbf{A}_j$$

We can observe from equation (4.11) that only one term in the equation, $\boldsymbol{\Pi}\mathbf{y}_{t-1}$, is in levels, cointegration relations depend crucially on the property of matrix $\boldsymbol{\Pi}$. It is clear that $\boldsymbol{\Pi}\mathbf{y}_{t-1}$ must be either $I(0)$ or zero except that $\mathbf{y}_t$ is already stationary. There are three situations:

(a) $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ has a reduced rank $0 < r < k$,
(b) $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ has a rank of zero, and
(c) $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ has a full rank.

Under situation (a), $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are both $k \times r$ matrices and have a rank of r. There are r cointegration vectors $\boldsymbol{\beta}'\mathbf{y}_t$ which are stationary $I(0)$ series. It is equivalent to having r common trends among $\mathbf{y}_t$. The stationarity of $\boldsymbol{\beta}'\mathbf{y}_t$ implies a long-run relationship among $\mathbf{y}_t$ or a sub-set of $\mathbf{y}_t$ – the variables in the cointegration vectors will not depart from each other over time. $\boldsymbol{\beta}'\mathbf{y}_t$ are also error correction terms in that departure of individual variables in the cointegration vectors from the equilibrium will be subsequently reversed back to the equilibrium – a dynamic adjustment process called error correction mechanism (ECM). Equation (4.11) is therefore called VAR with ECM. Under situation (b), there is no cointegration relation among $\mathbf{y}_t$ and the variables in levels do not enter equation (4.11) and equation (4.11) becomes a simple VAR without ECM. The variables in levels are already stationary under situation (c).

Depending on whether $\mathbf{y}_t$ and/or cointegration vectors have an intercept and/or deterministic trend, there are five models in practice: (a) there are no deterministic trends in $\mathbf{y}_t$ and no intercepts in cointegration vectors; (b) there is no deterministic trend in $\mathbf{y}_t$ but there are intercepts in cointegration vectors; (c) there are deterministic trends in $\mathbf{y}_t$ and intercepts in cointegration vectors; (d) there are deterministic trends in $\mathbf{y}_t$ and in cointegration vectors; (e) there are quadratic trends in $\mathbf{y}_t$ and deterministic trends in cointegration vectors. For details

of these specifications, see Johansen and Juselius (1990), and for the critical values of test statistics see Osterwald-Lenum (1992). The Johansen test is a kind of principal component analysis where eigenvalues of $\boldsymbol{\Pi}$ are calculated through a maximisation procedure. Then, five specifications or hypotheses are tested using the maximum eigenvalue statistic and the trace statistic which often convey contradictory messages. To test the hypothesis that there are r cointegration vectors against the alternative of $(r+1)$ cointegration vectors, there is the following maximum eigenvalue statistic:

$$\lambda_{\max} = -T \ln(1 - \hat{\lambda}_{r+1}) \tag{4.12}$$

where $\hat{\lambda}_r$ is the eigenvalue corresponding to r cointegration vectors and T is the number of observations. The trace statistic is calculated as:

$$\lambda_{\text{trace}} = -T \sum_{i=r+1}^{k} \ln(1 - \hat{\lambda}_i) \tag{4.13}$$

Indeed, the trace statistic for the existence of r cointegration vectors is the sum of the maximum eigenvalue statistics for from zero up to r cointegration vectors.

4.3. Common trends and common cycles

It should be noted that cointegration is not exactly the same as common trend analysis. While cointegration implies common trends it also requires non-existence of uncommon trends. A group of time series variables can share one or more common trends but the variables are not cointegrated because, for example, one of the variables, y_{2t}, also possesses, in addition to the common trends, a trend which is unique to itself and uncommon to others. Under such circumstances, the cointegration vector $\boldsymbol{\beta}' \mathbf{y}_t$ in equation (4.11) will exclude y_{2t} and it appears that y_{2t} does not share common trends with other variables in $\mathbf{y}_t$. Consider the following k-variable system:

$$
\begin{aligned}
y_{1t} &= a_{11} T_{1t} + \cdots + a_{1r} T_{rt} + \tau_{1t} + c_{1t} + \varepsilon_{1t} \\
y_{2t} &= a_{21} T_{1t} + \cdots + a_{2r} T_{rt} + \tau_{2t} + c_{2t} + \varepsilon_{2t} \\
&\cdots \\
y_{kt} &= a_{k1} T_{1t} + \cdots + a_{kr} T_{rt} + \tau_{kt} + c_{kt} + \varepsilon_{kt}
\end{aligned}
\tag{4.14}
$$

where $T_{it}, i = 1, \ldots r$ is the ith common trend, $\tau_{jt}, j = 1, \ldots k$ is the unique trend in y_{jt}, and $c_{jt}, j = 1, \ldots k$ is the cycle or stationary component in y_{jt}. If there are no unique trends, i.e. $\tau_{jt} = 0, j = 1, \ldots k$, then from linear algebra we know that a certain linear combination of $y_{jt}, j = 1, \ldots k$ is zero when $r < k$. So there are only cycles or stationary components, $c_{jt}, j = 1, \ldots k$, left in the linear combination of $y_{jt}, j = 1, \ldots k$, which exhibits no trends. This is exactly the idea of cointegration

discussed above. When there is unique trend in, for example, y_{2t} (i.e. $\tau_{2t} \neq 0$; $\tau_{jt} = 0, j \neq 2$), y_{2t} will not be cointegrated with any other variables in the system as any linear combination involving y_{2t} will be non-stationary, though y_{2t} does share common trends with the rest of the variables. It is clear that if y_{2t} does join other variables in $\boldsymbol{\beta}' \mathbf{y}_t$, it must contain no unique trend. For convenience, common trends are treated as the same as cointegration in this chapter. That is, unique trends are excluded from analysis.

In the following, we extend cointegration and common trend analysis to the case of cycles. It is said (Engle and Kozicki, 1993; Vahid and Engle, 1993; Engle and Issler, 1995) there are common cycles (in the same spirit, uncommon cycles are excluded from analysis) among $\mathbf{y}_t$ in equation (4.10) if there exists a vector $\tilde{\boldsymbol{\beta}}$ such that:

$$\tilde{\boldsymbol{\beta}}' \mathbf{y}_t = \tilde{\boldsymbol{\beta}}' \boldsymbol{\varepsilon}_t \tag{4.15}$$

That is, a combination of the time series in $\mathbf{y}_t$ exhibits no cyclical movement or fluctuation. Common trends and common cycles are two major common factors driving economic and financial variables to move and develop in a related way.[3] It is therefore helpful to inspect them together in a unified dynamic system.

According to the Wold representation theorem, time series or a vector of time series can be expressed as an infinite moving average process:

$$\Delta \mathbf{y}_t = \mathbf{C}(L) \boldsymbol{\varepsilon}_t, \qquad \mathbf{C}(L) = \mathbf{I} + \mathbf{C}_1 L + \mathbf{C}_2 L^2 + \cdots \tag{4.16}$$

$\mathbf{C}(L)$ can be decomposed as $\mathbf{C}(1) + (1 - L)\mathbf{C}^*(L)$, therefore:

$$\Delta \mathbf{y}_t = \mathbf{C}(1) \boldsymbol{\varepsilon}_t + (1 - L)\mathbf{C}^*(L) \boldsymbol{\varepsilon}_t, \qquad \mathbf{C}_i^* = \sum_{j>i} -\mathbf{C}_j, \qquad \mathbf{C}_0^* = \mathbf{I} - \mathbf{C}(1) \tag{4.17}$$

Taking the summation to get the variables in levels:

$$\mathbf{y}_t = \mathbf{C}(1) \sum_{i=0}^{\infty} \boldsymbol{\varepsilon}_{t-i} + \mathbf{C}^*(L) \boldsymbol{\varepsilon}_t \tag{4.18}$$

Equation (4.18) is the Stock and Watson (1988) multivariate generalisation of the Beveridge and Nelson (1981) trend-cycle decomposition and is referred to as the BNSW decomposition. Common trends in the sense of cointegration require:

$$\boldsymbol{\beta}' \mathbf{C}(1) = 0 \tag{4.19}$$

and common cycles require:

$$\tilde{\boldsymbol{\beta}}' \mathbf{C}^*(L) = 0 \tag{4.20}$$

Equation (4.18) can be written as the sum of two components of trends and cycles:

$$\mathbf{y}_t = \mathbf{T}_t + \mathbf{C}_t \tag{4.21}$$

When the sum of the rank of $\boldsymbol{\beta}$, $\mathbf{r}$, and the rank of $\tilde{\boldsymbol{\beta}}$, $\mathbf{s}$, is equal to k, the stack of $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$ is a $k \times k$ full rank matrix:

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta} & \tilde{\boldsymbol{\beta}} \end{bmatrix} \tag{4.22}$$

and trends and cycles can be exclusively expressed in the common factor coefficient vectors, $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$, and their combinations. According to equations (4.19) and (4.20):

$$\begin{bmatrix} \boldsymbol{\beta} & \tilde{\boldsymbol{\beta}} \end{bmatrix}' \mathbf{y}_t = \begin{bmatrix} \boldsymbol{\beta}' \mathbf{y}_t \\ \tilde{\boldsymbol{\beta}}' \mathbf{y}_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\beta}' \mathbf{C}^*(\mathbf{L}) \boldsymbol{\varepsilon}_t \\ \tilde{\boldsymbol{\beta}}' \mathbf{C}(1) \sum_{i=0}^{\infty} \boldsymbol{\varepsilon}_{t-i} \end{bmatrix}$$

So:

$$\mathbf{y}_t = \begin{bmatrix} \boldsymbol{\beta}' \\ \tilde{\boldsymbol{\beta}}' \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\beta}' \mathbf{C}^*(\mathbf{L}) \boldsymbol{\varepsilon}_t \\ \tilde{\boldsymbol{\beta}}' \mathbf{C}(1) \sum_{i=0}^{\infty} \boldsymbol{\varepsilon}_{t-i} \end{bmatrix} \tag{4.23}$$

Define $\mathbf{B}^{-1} = \begin{bmatrix} \boldsymbol{\beta}^{-1} & \tilde{\boldsymbol{\beta}}^{-1} \end{bmatrix}^4$ and refer to equation (4.18), we have:

$$\mathbf{y}_t = \begin{bmatrix} \boldsymbol{\beta}^{-1} & \tilde{\boldsymbol{\beta}}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}' \mathbf{C}^*(\mathbf{L}) \boldsymbol{\varepsilon}_t \\ \tilde{\boldsymbol{\beta}}' \mathbf{C}(1) \sum_{i=0}^{\infty} \boldsymbol{\varepsilon}_{t-i} \end{bmatrix} = \boldsymbol{\beta}^{-1} \boldsymbol{\beta}' \mathbf{C}^*(\mathbf{L}) \boldsymbol{\varepsilon}_t + \tilde{\boldsymbol{\beta}}^{-1} \tilde{\boldsymbol{\beta}}' \mathbf{C}(1) \sum_{i=0}^{\infty} \boldsymbol{\varepsilon}_{t-i}$$

$$= \boldsymbol{\beta}^{-1} \boldsymbol{\beta}' \mathbf{y}_t + \tilde{\boldsymbol{\beta}}^{-1} \tilde{\boldsymbol{\beta}}' \mathbf{y}_t = \mathbf{C}_t + \mathbf{T}_t \tag{4.24}$$

Therefore, we get $\mathbf{C}_t = \boldsymbol{\beta}^{-1} \boldsymbol{\beta}' \mathbf{y}_t$, and $\mathbf{T}_t = \tilde{\boldsymbol{\beta}}^{-1} \tilde{\boldsymbol{\beta}}' \mathbf{y}_t$, exclusively expressed in the common factor coefficient vectors, $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\beta}}$, and their combinations.

4.4. Examples and cases

It is probably not worthwhile demonstrating any unit root test examples individually nowadays since these tests have been made straightforwardly simple. Nevertheless, unit root tests are still routine procedures prior to cointegration analysis, i.e. studies of cointegration will almost inevitably involve unit root tests. Accordingly, one case on cointegration and one case on common cycles are presented in the following which largely cover the topics of this chapter.

Example 4.1

This is a case on dynamic links and interactions between American Depository Receipts (ADRs) and their underlying foreign stocks by Kim *et al.* (2000). ADRs are certificates issued by a US bank which represent

Continued

---

indirect ownership of a certain number of shares in a specific foreign firm. Shares are held on deposit in the firm's home bank. ADRs are traded in US dollars and investors receive dividends in US dollars too. Therefore, returns on ADRs reflect the domestic returns on the stock as well as the exchange rate effect. ADRs have become popular in the US due to their diversification benefits, especially when US investors have little knowledge in foreign countries' business and political systems, and risks associated with investing in foreign securities may be overestimated.

In addition to the factors of underlying foreign stocks and the exchange rate, the paper has also considered the influence of the US stock market on ADRs returns. To this end, they use the cointegration approach and other models to examine the effect on ADR of the three factors. Their results show that the price of the underlying stock is most important, whereas the exchange rate and the US market also have an impact on ADR prices. We only present results related to cointegration analysis of 21 British firms. The data set used in the paper is daily closing prices from 4 January 1988 to 31 December 1991.

The first thing to do prior to cointegration tests is almost a routine check on whether there is a unit root in the time series data, as we require $I(1)$ series to carry out cointegration analysis. The paper adopts the ADF test to examine the existence of a unit root, with the critical values being taken from Davidson and Mackinnon (1993), obtained from a much larger set of simulations than those tabulated by Dickey and Fuller. The lag length in the ADF test is chosen such that the Q-statistic at 36 lags indicates no serial correlation in the residuals. The lag length can also be chosen by using the Akaike information criterion (AIC) or Schwarz criterion (SC), or more ideally, a combination of the Q-statistic and one of the AIC or SC which, though, may produce non-conciliatory recommendations. It can be seen from Table 4.1 that all the series, except series 8 and 19 (interestingly both ADRs and underlying stocks), have a unit root in levels and no unit root in the first difference. Although the null hypothesis of a unit root is rejected for series 8 and 19 in levels, the rejection is at a rather low 10 per cent significance level. So, all the series are treated as $I(0)$ and cointegration analysis can be carried out for all of them. In Table 4.2, the exchange rate and one of the US stock market indices, S&P500, are also confirmed to be $I(1)$ series, and can be included in cointegration analysis as well.

The results of cointegration analysis between ADRs, corresponding foreign stocks, the exchange rate and the S&P500 index, are reported in Table 4.3. The lag length k is chosen by Sims' likelihood ratio test. Both trace and eigenvalue test statistics indicate that for all 21 groups, there exists at least one cointegrating relationship among the variables. Nine groups have at least two and three groups have at least three cointegrating relationships.

Table 4.1 Augmented Dickey–Fuller unit root tests – ADRs and underlying foreign stocks, UK

Firm	ADR		Underlying	
	Level	*First difference*	*Level*	*First difference*
1	−2.559	−35.842***	−2.461	−22.132***
2	−1.245	−13.821***	−1.725	−19.753***
3	−1.652	−14.873***	−1.823	−12.694***
4	−2.235	−15.985***	−1.927	−13.346***
5	−1.985	−26.879***	−1.969	−28.566***
6	−2.237	−27.522***	−1.878	−25.997***
7	−2.334	−20.464***	−1.200	−23.489***
8	−2.652*	−30.435***	−2.800*	−29.833***
9	−1.287	−10.156***	−2.382	−14.489***
10	−1.823	−26.372***	−1.014	−21.788***
11	−1.021	−27.825***	−1.087	−19.482***
12	−1.934	−29.225***	−2.425	−27.125***
13	−2.324	−13.223***	−1.894	−12.854***
14	−1.997	−17.325***	−1.823	−16.478***
15	−1.333	−11.528***	−1.458	−37.311***
16	−1.223	−10.285***	−1.253	−18.244***
17	−1.110	−16.742***	−2.182	−33.245***
18	−1.559	−14.522***	−1.285	−17.354***
19	−2.678*	−22.485***	−2.677*	−15.660***
20	−1.546	−14.266***	−1.024	−14.266***
21	−2.364	−22.333***	−1.625	−24.757***

Asymptotic critical values are from Davidson and Mackinnon (1993). Lag length *K* is chosen such that the *Q*-statistic at 36 lags indicates absence of autocorrelation in the residuals. Estimation period is 4 January 1988–31 December 1991.
*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.

Table 4.2 Augmented Dickey–Fuller unit root tests – the exchange rate and the S&P 500 index

	Level	*First difference*
£ vis-à-vis $	−1.625	−19.124***
S&P 500 index	−2.115	−20.254***

Asymptotic critical values are from Davidson and Mackinnon (1993). Lag length *K* is chosen such that the *Q*-statistic at 36 lags indicates absence of autocorrelation in the residuals. Estimation period is 4 January 1988–31 December 1991.
*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.

Continued

Table 4.3 Johansen multivariate cointegration tests – United Kingdom

Firm	Trace				λ_{max}			
	$r = 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$	$r = 0$	$r \leq 1$	$r \leq 2$	$r \leq 3$
1	91.22***	21.28	2.95	1.04	70.11***	10.09	1.89	1.04
2	52.18***	12.45	6.06	1.25	36.37***	10.65	4.75	1.25
3	68.02***	20.24	5.79	0.50	84.67***	12.34	5.00	0.50
4	105.24***	28.45*	7.64	1.25	96.77***	19.65*	6.75	1.25
5	45.33*	11.02	1.24	0.37	39.32***	9.88	1.05	0.37
6	163.26***	32.84**	9.75	3.94	141.21***	27.87***	13.25**	3.94
7	85.24***	19.45	2.25	1.00	66.47***	10.65	1.75	1.00
8	150.33***	30.02*	8.24	3.54	120.32***	20.78*	6.45	3.54
9	49.23**	12.02	1.29	0.98	29.32**	9.78	9.24	0.98
10	50.24**	13.45	1.54	1.08	46.37***	10.65	3.25	1.08
11	190.33***	38.02***	18.24***	3.99	145.31***	28.88**	11.48	3.99
12	96.96***	21.84	3.00	1.52	72.50***	12.09	2.69	1.52
13	150.24***	30.00*	7.34	3.24	120.22***	20.74*	5.45	3.24
14	199.43***	42.02***	19.24***	4.57	150.32***	38.99***	18.45***	4.57
15	153.33***	31.25**	8.66	3.25	125.43***	21.27**	5.75	3.25
16	81.43***	21.34	5.24	2.08	52.45***	17.24	3.78	2.08
17	210.24***	68.24***	21.78***	4.02	139.32***	34.28***	16.27**	4.02
18	62.96***	13.11	1.75	0.99	42.11***	10.09	1.29	0.99
19	49.24**	9.92	1.24	0.61	27.88**	8.45	1.05	0.61
20	120.33***	24.91	6.24	2.01	84.56***	15.74	3.45	2.01
21	173.86***	33.24**	8.03	4.06	121.54***	33.34***	10.49	4.06

The cointegration equation is based on four variables: (1) British ADRs, (2) British underlying shares, (3) British pound spot exchange rates, and (4) the S&P 500 index cash prices. Estimation period is 4 January 1988–31 December 1991.
*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.

Each group's cointegrating vector is calculated and incorporated in the VAR to form a VAR–ECM model. Based on the estimated VAR–ECM model, the paper has further performed variance decomposition and impulse response analysis which are beyond the reach of this chapter.[5] While cointegration analysis indicate a dynamic adjustment process and long-run equilibrium relationship among ADRs and the three factors, results from variance decomposition and impulse responses suggest that the largest effect on ADRs is due to shocks in their underlying stocks. Nevertheless, the exchange rate also has a role and that role is growing in recent years. The effect of the US stock market has been found but the effect is small. More likely, the last link might be superficial and due to a common factor driving both the US and foreign markets or the US stock market and the foreign exchange market.

Example 4.2

This is an example mainly on common cycles, but also covers common trends, among annual sectoral per-capita real GNP of the US economy 1947–1989, from a paper entitled 'Estimating common sectoral cycles' by Engle and Issler (1995). The sectors examined are agriculture, forestry and fisheries, mining, manufacturing, construction, wholesale and retail, transportation and public utilities, finance, insurance and real estate, and services.

The paper does not check for unit roots itself; instead it cites the results of Durlauf (1989) that the sectoral GNP data are $I(1)$ processes. The first set of empirical results is on cointegration or common trends, and the second set of results is common cycles. Table 4.4 reports the cointegration results which show that there are two cointegration vectors judged by both the trace and maximum eigenvalue test statistics, adopting the model with unrestricted intercept and a linear time trend. The trace statistic also points to a third coinegration relation at a low significance level. The paper then sets up a VAR–ECM model of two cointegration vectors to investigate the dynamics among the sectors.

The common cycle test is based on canonical correlation[6] and the results are reported in Table 4.5. They are interpreted in this way: the number of common cycle relations is the number of zero canonical correlations. Since the test statistic (second column) rejects that five canonical correlations are zero and cannot reject that six or more canonical correlations are zero, the number of common cycle relations is decided to be six. To find a larger number of common factor relations must be rather confusing. Adeptly, the paper suggests that very similar cyclical behaviour for sectors be observed without going into detail of these common cycle vector coefficients.

Table 4.4 Cointegration results – Johansen's approach (1988)

No of CI vectors	λ_{max}	5% critical value	Trace	5% critical value
At most 7	1.9	3.7	1.9	3.7
At most 6	10.5	16.9	12.4	18.2
At most 5	15.5	23.8	27.6	34.5
At most 4	20.3	30.3	47.9	54.6
At most 3	25.3	36.4	73.2	77.7
At most 2	32.9	42.5	106.1*	104.9
At most 1	68.4**	48.4	174.5**	136.6
At most 0	108.4**	54.2	282.8**	170.8

*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.

Continued

Table 4.5 Common cycle results

No of common cycles	Squared canonical correlation ρ_i^2	Pr > F
8 ($\rho_i^2 = 0, i = 1, \ldots 8$)	0.9674	0.0001
7 ($\rho_i^2 = 0, i = 1, \ldots 7$)	0.8949	0.0113
6 ($\rho_i^2 = 0, i = 1, \ldots 6$)	0.7464	0.4198
5 ($\rho_i^2 = 0, i = 1, \ldots 5$)	0.5855	0.7237
4 ($\rho_i^2 = 0, i = 1, \ldots 4$)	0.5130	0.7842
3 ($\rho_i^2 = 0, i = 1, \ldots 3$)	0.4367	0.8088
2 ($\rho_i^2 = 0, i = 1, 2$)	0.3876	0.7922
1 ($\rho_1^2 = 0$)	0.2775	0.7848

Because the sum of the number of cointegration or common trend relations and the number of common cycle relations is eight – the number of sectors or variables in the VAR, trends and cycles can be exclusively expressed in the common factor coefficient vectors and their combinations. Table 4.5 presents these vectors – six of them are common cycle coefficient vectors and two of them common trend coefficient vectors.

4.5. Empirical literature

Research on unit roots and tests for stationarity is one of the frontiers in contemporary time series econometrics. The distinction between stationary and non-stationary time series data can, explicitly or implicitly, reflect the data's economic or financial characteristics and attributes. For example, if a variable's current state or value is derived through accumulation of all previous increases (decreases as negative increases) in its value, then this variable is almost certainly non-stationary. If a variable is a relative measure, e.g. the growth rate in GDP, or the rate of return on a stock, which has nothing to do with its history, then it is more likely to be stationary, though non-stationarity cannot be ruled out when there is non-trivial change in the rate (acceleration). For some other relative measures, such as dividend yields (dividend/price), the percentage of public sector borrowing requirement in GDP (PSBR/GDP), or asset turnover (sales/asset value), it is an empirical matter whether the time series data are stationary or not. Indeed, stationarity of this type of relative measures amounts to cointegration, with the coinegration vector being restricted to $[1, -1]$, between the two variables involved in the construction of the measure (when logarithm is taken). So, we derive the concept of cointegration, another frontier in time series econometrics, in a very natural way, closely related to real world phenomena. Extending this relative measure to cross-sections, e.g. data of different entities, we have cointegration in general forms, to examine whether these entities progress in pace or proportionally in the long-run. As a result, tests for unit roots and cointegration infer the

attributes of economic and financial variables and their relationships reflected by the characteristics of time series data. To a lesser significant extent, there is research in common cycles that two or more variables, which are stationary, move in a rather similar way in the short-term.

Much research on the subjects focuses on both economic analysis and financial markets, in a variety of application areas. So let us start with the interest rate and the exchange rate which are to the common interest of most economic and financial variables. Examining well known international parity conditions of Covered Interest Parity (CIP), Uncovered Interest Parity (UIP), the Forward Rate Hypothesis (FRH), Purchasing Power Parity (PPP), and the International Fisher Effect (IFE), Turtle and Abeysekera (1996) adopt cointegration procedures to test the validity of these hypotheses implied by the cointegration relationship between spot rates, forward rates, interest rates, and inflation rates using monthly data from January 1975 to August 1990 for Canada, Germany, Japan, and the UK against the US. They claim that the cointegration test results generally support the relationships considered. In a more focused study, MacDonald and Nagayasu (2000) investigate the long-run relationship between real exchange rates and real interest rate differentials with a panel data set consisting of 14 industrialised countries, over the recent floating period. Similar to a few of other empirical studies with panel data, the procedure of panel unit root and cointegration tests tends to favour stationarity with which the paper finds evidence of statistically significant long-run relationships between real exchange rates and real interest rate differentials. Likewise, Wu and Fountas (2000) suggest bilateral real interest rate convergence between the US and the rest of the G7 countries, and Felmingham *et al.* (2000) find interdependence between the Australian short-term real interest rates and those of the US, Japan, the UK, Canada, Germany, and New Zealand during 1970 and 1997, after accommodating regime shifts in the time series. Fountas and Wu (1999) show similar findings of real interest rate convergence among European monetary mechanism countries for the period of 1979–1993. Chiang and Kim (2000) present a set of empirical results for Eurocurrency market rates. They find the domestic short-term interest rate is cointegrated with longer-term interest rates of a particular country; and the domestic short-term interest rate is also cointegrated with the comparable foreign short-term interest rate adjusted for the foreign exchange forward premium/discount. They consequently set up an error correction model including both cointegration vectors, and claim that the model has improvements in explaining short-term interest rate movements. Extending research in foreign exchange rates to a non-standard setting, Siddiki (2000) examines the determinants of black market exchange rates in India using annual data from 1955 to 1994 in the framework of unit root and cointegration analysis. The paper confirms that the import capacity of official foreign exchange reserves and restrictions on international trade are two important determinants of black market rates in India, and finds that black market rates are negatively affected by a low level of official foreign exchange reserves and positively affected by a high level of trade restrictions, as well as interest rate policies. Of more practical orientation is a study by Darrat *et al.* (1998) on the possible link between the

mortgage loan rate and the deposit rate, and the question of which rate leads the other. While the deposit-cost mark-up theory suggests that the cost of attracting funds (deposit rates) determines prices (mortgage loan rates), mortgage loan rates may induce changes in deposit interest rates in a mechanism of the reverse chain of events. The authors employ cointegration and Granger causality tests to empirically examine these alternative hypotheses, using monthly data over the period of 1970 to 1994. The results appear to accommodate both hypotheses, that there exists a bi-directional causality between mortgage loan rates and deposit interest rates in an error correction model where the two types of rates exhibit a cointegration relationship. Many recent studies of the kind can be found, for example, in Toma (1999), Wright (2000), Cheng (1999), Pesaran *et al.* (2000), and Koustas and Serletis (1999).

Research on long-run relationships in stock markets is controversial in that it constitutes a contest to market efficiency. Adopting a pragmatic stance in empirical analysis, Harasty and Roulet (2000) employ the Engle–Granger two-step method for cointegration analysis and error correction modelling of stock market movements in 17 countries. They present in- and out-of-sample tests of the model's ability to forecast future stock market returns, and their results, it is claimed, indicate that the error correction model does have predictive power and can thus be a useful tool in the investment decision process. A long-run cointegration relationship has also been found to exist in Eastern European stock markets between 1995 and 1997 by Jochum *et al.* (1999). They report that the cointegration relationship has disappeared after the 1977 stock market crisis. With a total sample period of three years and the post-crisis sub-period of only one year, these results can hardly be of helpful implications, though the problem is mainly due to the availability of data. Olienyk *et al.* (1999) attempt to avoid the problems of non-synchronous trading, fluctuations in foreign exchange rates, non-liquidity, trading restrictions, and index replication by using World Equity Benchmark Shares (WEBS) to effectively represent the world's stock markets. They observe that long-run relationship exists among the 18 market indices, as well as between individual closed-end country funds and their own country's WEBS. They further find that there exists short-term Granger causality between these series, implying market inefficiencies and short-term arbitrage opportunities. In an effort to explain market efficiency in the context of cointegration, Hassapis *et al.* (1999) extend the work by Dwyer and Wallace (1992) through investigating the linkages among international commodity markets in the long-run and the short-term. Efficiency in these markets requires that the corresponding real exchange rates be martingales with respect to any information set available in the public domain. In a VAR consisting only of real exchange rates, it is shown that necessary and sufficient conditions for joint efficiency of all the markets under consideration amount to the VAR being of order one (Markovness) and non-cointegrated. On the contrary, in a VAR extended by other potentially 'relevant' variables, such as the corresponding real interest rates, non-cointegration and Markovness are only sufficient conditions for the same commodity markets to be characterised as jointly efficient.

In labour market studies, the relationships between wage costs and employment have been subject to extensive scrutiny for many decades. The new techniques of unit root tests and cointegration offer an additional dimension to the research in terms of the long-run characteristics of wages and employment and the long-run relationship between wages and employment. In this framework, Bender and Theodossiou (1999) investigate the relationship between employment and real wages for ten countries since 1950. Their results suggest that there is little evidence of cointegration between real wages and employment and consequently reject the neoclassical hypothesis of a long-run relationship between these two important variables. Including more variables in the analysis of the Mexican labour market, Lopez (1999) finds cointegration relationships between employment and output, and among nominal wages, minimum wages, the price index, and labour productivity. The results do not directly contradict those of Bender and Theodossiou (1999) but they offer explanations to the dynamic adjustment of employment and wages to a set of macroeconomic variables. Similarly, Carstensen and Hansen (2000) find two common trends, which push unemployment, in the West German labour market with a structural VAR incorporating cointegration.

Various other recent application examples cover the examination of the Fisher effect by Koustas and Serletis (1999) in Belgium, Canada, Denmark, France, Germany, Greece, Ireland, Japan, the Netherlands, the UK and the US with results generally rejecting the Fisher hypothesis, and by Malliaropulos (2000) for the US who supports the hypothesis; interactions between the stock market and macroeconomic variables by Choi *et al.* (1999) who suggest that stock markets help predict industrial production in the US, UK, Japan and Canada out of G-7, and by Nasseh and Strauss (2000) where not only domestic, but also international, macroeconomic variables, enter the cointegration vectors to share long-run relationships with stock prices; long-run relationships between real estate as represented by REITs, and the bond market and stock market by Glascock *et al.* (2000); and joint efficiency of the US stock market and foreign exchange markets by Rapp *et al.* (1999). It is indeed a very long list but yet to exhaust all the studies in these areas.

Notes

1 Readers familiar with difference equations, deterministic and/or stochastic, would understand this easily. Equation (4.6) also has a pole at $p = \infty$, which is not as important in relation to the topic.
2 Precisely, it is the spectrum obtained from letting Δy_t pass a rectangular window of size M.
3 Other common factors include regime shifts, see, for example, co-break in Hendry and Mizon (1998).
4 Notice $\boldsymbol{\beta}^{-1}$ is not the inverse matrix of $\boldsymbol{\beta}$ (such inverse matrix does not exist), it is simply the first r columns of $\mathbf{B}^{-1} = \begin{bmatrix} \boldsymbol{\beta}' \\ \tilde{\boldsymbol{\beta}}' \end{bmatrix}^{-1}$, similarly $\tilde{\boldsymbol{\beta}}^{-1}$ is the last s columns of $\mathbf{B}^{-1}$.
5 We can introduce briefly the ideas of variance decomposition and impulse response here. Variance decomposition is to inspect the contributions to one sector's variance from all

other sectors, including itself, so the relative importance of these sectors can be evaluated. Impulse response analysis is to examine the impact of a unit shock in one sector on the other; similar to variance decomposition, the influence of one sector on the other and the relative importance of all the sectors to an individual sector can be evaluated. Both impulse response and variance decomposition, especially the former, are usually carried out over a long time horizon; and impulse response is normally presented in the form of visual graphs.

6 A technique similar to, if all appropriate, Johansen's multivariate cointegration analysis and is its stationary counterpart. The technique is not widely applied as more than one common cycle relation, similar to more than one cointegration relation, is difficult to be conferred a meaningful economic interpretation. If feasible, pair wise analysis will usually be applied.

Questions and problems

1 Discuss the concept of stationarity and non-stationarity in relation to the characteristics of financial variables, e.g. prices and returns are the accumulation of income (dividends) over time, so are their statistical properties.

2 Describe a unit root process and show it does not have a constant limited variance.

3 Discuss the cointegration relationship in econometrics and the comovement of certain non-stationary financial and economic variables, e.g. dividends and prices, inflation and nominal interest rates, and industrial production and stock market returns.

4 What are the features of common cycles in contrast to common trends and cointegration?

5 Discuss the common cycle relationship in econometrics and the comovement of certain stationary variables in economics and finance.

6 Discuss in what circumstances cointegration implies market inefficiency and in what circumstances cointegration means market efficiency.

7 Collect data from Datastream to test for unit roots in the following time series:

   (a) GDP of the UK, US, Japan, China, Russia, and Brazil in logarithms,
   (b) total return series of IBM, Microsoft, Sage, Motorola, Intel, Vodafone, and Telefonica in logarithms,
   (c) nominal interests in selected countries.

   What do you find of their characteristics?

8 Test for unit roots in the above time series in log differences. What do you find of their characteristics?

9 Collect data from Datastream to test for cointegration between the following pairs:

   (a) the sterling *vis-à-vis* US$ exchange rates, spot and 30 days forward,
   (b) Tesco and Sainsbury's share prices,
   (c) UK underlying RPI and the Bank of England base rate.

   Discuss your findings.

References

Bender, K.A. and Theodossiou, I. (1999), International comparisons of the real wage–employment relationship, *Journal of Post Keynesian Economics*, **21**, 621–637.

Beveridge, S. and Nelson, C.R. (1981), A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycles', *Journal of Monetary Economics*, **7**, 151–174.

Carstensen, K. and Hansen, G. (2000), Cointegration and common trends on the West German labour market, *Empirical Economics*, **25**, 475–493.

Cheng, B.S. (1999), Beyond the purchasing power parity: testing for cointegration and causality between exchange rates, prices, and interest rates, *Journal of International Money and Finance*, **18**, 911–924.

Chiang, T.C. and Kim, D. (2000), Short-term Eurocurrency rate behavior and specifications of cointegrating processes, *International Review of Economics and Finance*, **9**, 157–179.

Choi, J.J., Hauser, S. and Kopecky, K.J. (1999), Does the stock market predict real activity? Time series evidence from the G-7 countries, *Journal of Banking and Finance*, **23**, 1771–1792.

Darrat, A.F., Dickens, R.N. and Glascock, J.L. (1998), Mortgage loan rates and deposit costs: are they reliably linked? *Journal of Real Estate Finance and Economics*, **16**, 27–42.

Davidson, R. and MacKinnon, J.G. (1993), *Estimation and Inference in Econometrics*, Oxford University Press, England.

Dickey, D.A. and Fuller, W.A. (1979), Distribution of the estimators for autoregressive time series with a unit root, *Journal of the American Statistical Association*, **74**, 427–431.

Dickey, D.A. and Fuller, W.A. (1981), The likelihood ratio statistics for autoregressive time series with a unit root, *Econometrica*, **49**, 1057–1072.

Durlauf, S.N. (1989), Output persistence, economic structure, and the choice of stabilisation policy, *Brookings Papers on Economic Activity*, **2**, 69–136.

Dwyer, G.P. Jr. and Wallace, M.S. (1992), Cointegration and market efficiency, *Journal of International Money and Finance*, **11**, 318–327.

Engle, R.F. and Granger, C.W.J. (1987), Co-integration and error correction: Representation, estimation, and testing, *Econometrica*, **55**, 251–267.

Engle, R.F. and Issler, J.V. (1995), Estimating common sectoral cycles, *Journal of Monetary Economics*, **35**, 83–113.

Engle, R.F. and Kozicki, S. (1993), Testing for common features, *Journal of Business and Economic Statistics*, **11**, 369–395.

Felmingham, B., Qing, Z. and Healy, T. (2000), The interdependence of Australian and foreign real interest rates, *Economic Record*, **76**, 163–171.

Fountas, S. and Wu, J.L. (1999), Testing for real interest rate convergence in European countries, *Scottish Journal of Political Economy*, **46**, 158–174.

Glascock, J.L., Lu, C. and So, R.W. (2000), Further evidence on the integration of REIT, bond, and stock returns, *Journal of Real Estate Finance and Economics*, **20**, 177–194.

Harasty, H. and Roulet, J. (2000), Modelling stock market returns, *Journal of Portfolio Management*, **26**(2), 33–46.

Hassapis, C., Kalyvitis, S.C. and Pittis, N. (1999), Cointegration and joint efficiency of international commodity markets, *Quarterly Review of Economics and Finance*, **39**, 213–231.

Hendry, D.F. and Mizon, G.E. (1998), Exogeneity, causality, and co-breaking in economic policy analysis of a small econometric model of money in the UK, *Empirical Economics*, **23**, 267–294.

Im, K.S., Pesaran, M.H. and Shin, Y. (1995), Testing for unit roots in heterogeneous panels, *University of Cambridge, Department of Applied Economics Working Paper, Amalgamated Series*: 9526.

Jochum, C., Kirchgassner, G. and Platek, M. (1999), A long-run relationship between Eastern European stock markets? Cointegration and the 1997/98 crisis in emerging markets, *Weltwirtschaftliches Archiv – Review of World Economics*, **135**, 454–479.

Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, **12**, 231–254.

Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica*, **59**, 1551–1580.

Johansen, S. and Juselius, K. (1990), Maximum likelihood estimation and inference on cointegration – with applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, **52**, 169–210.

Kao, C. and Chiang, M.H. (1998), On the estimation and inference of a cointegrated regression in panel data, *Centre for Policy Research Working Paper, Syracuse University*.

Kim, M., Szakmary, A.C. and Mathur, I. (2000), Price transmission dynamics between ADRs and their underlying foreign securities, *Journal of Banking and Finance*, **24**, 1359–1382.

Koustas, Z. and Serletis, A. (1999), On the Fisher effect, *Journal of Monetary Economics*, **44**, 105–130.

Kwiatkowski, D., Phillips, P.C.B, Schmidt, P. and Shin, Y. (1992), Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, **54**, 159–178.

Levin, A. and Lin, C.F. (1992), Unit root tests in panel data: asymptotic and finite sample properties, *University of California, San Diego Department of Economics Working Paper*: 92–23.

Levin, A. and Lin, C.F. (1993), Unit root tests in panel data: new results, *University of California, San Diego Department of Economics Working Paper*: 93–56.

Lopez, G.J. (1999), The macroeconomics of employment and wages in Mexico, *Labour*, **13**, 859–878.

MacDonald, R. and Nagayasu, J. (2000), The long-run relationship between real exchange rates and real interest rate differentials: a panel study, *IMF Staff Papers*, **47**, 116–128.

Maddala, G.S. and Wu, S. (1999), A comparative study of unit root tests with panel data and a new simple test, *Oxford Bulletin of Economics and Statistics*, **61**(0) (special issue), 631–652.

Malliaropulos, D. (2000), A note on nonstationarity, structural breaks, and the Fisher effect, *Journal of Banking and Finance*, **24**, 695–707.

Moon, H.R and Phillips, P.C.B (1999), Maximum likelihood estimation in panels with incidental trends, *Oxford Bulletin of Economics and Statistics*, **60**(0) (special issue), 711–747.

Nasseh, A. and Strauss, J. (2000), Stock prices and domestic and international macroeconomic activity: a cointegration approach, *Quarterly Review of Economics and Finance*, **40**, 229–245.

Olienyk, J.P., Schwebach, R.G. and Zumwalt, J.K. (1999), WEBS, SPDRs, and country funds: an analysis of international cointegration, *Journal of Multinational Financial Management*, **9**, 217–232.

Osterwald-Lenum, M. (1992), A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics, *Oxford Bulletin of Economics and Statistics*, **54**, 461–472.

Pedroni, P. (1999), Critical values for cointegration tests in heterogeneous panels with multiple regressors, *Oxford Bulletin of Economics and Statistics*, **61**(0) (special issue), 653–670.

Pesaran, M.H., Shin, Y. and Smith, R.J. (2000), Structural analysis of vector error correction models with exogenous I(1) variables, *Journal of Econometrics*, **97**, 293–343.

Phillips, P.C.B. and Perron. P. (1988), Testing for a unit root in time series regression, *Biometrika*, **75**, 335–346.

Rapp, T.A., Parker, M.E. and Phillips, M.D. (1999), An empirical investigation of the joint efficiency of the U.S. stock and foreign exchange markets, *Journal of Economics*, **25**, 63–71.

Siddiki, J.U. (2000), Black market exchange rates in India: an empirical analysis, *Empirical Economics*, **25**, 297–313.

Stock, J.H. and Watson, M.W. (1988), Testing for common trends, *Journal of the American Statistical Association*, **83**, 1097–1107.

Toma, M. (1999), A positive model of reserve requirements and interest on reserves: a clearinghouse interpretation of the Federal Reserve System, *Southern Economic Journal*, **66**, 101–116.

Turtle, H.J. and Abeysekera, S.P. (1996), An empirical examination of long run relationships in international markets, *Journal of Multinational Financial Management*, **6**, 109–134.

Vahid, F. and Engle, R.F. (1993), Common trends and common cycles, *Journal of Applied Econometrics*, **8**, 341–360.

Wright, G. (2000), Spot and period rates in the wet bulk shipping market: testing for long-run parity, *Journal of Transport Economics and Policy*, **34**, 291–300.

Wu, J.L. and Fountas, S. (2000), Real interest rate parity under regime shifts and implications for monetary policy, *The Manchester School of Economic and Social Studies*, **68**, 685–700.

5 Time-varying volatility models

GARCH and stochastic volatility

Time-varying volatility models have been popular since the early 1990s in empirical research in finance, following an influential paper 'Generalized Autoregressive Conditional Heteroskedasticity' by Bollerslev (1986). Models of this type are well known as GARCH in the time series econometrics literature. Time-varying volatility has been observed and documented in as early as 1982 (Engle, 1982) and was initially concerned with an economic phenomenon – time varying and autoregressive variance of inflation. Nevertheless, it was data availability and strong empirical research interest in finance, motivated by exploring any kind of market inefficiency, that encouraged the application and facilitated the development of these models and their variations. For instance, the GARCH in mean model is related to asset pricing with time-varying risk instead of constant risk in the traditional models such as the CAPM. An EGARCH (Exponential GARCH) model addresses asymmetry in volatility patterns which are well observed in corporate finance and financial markets and can sometimes be attributed to leverage effects. GARCH with t-distributions reflects fat tails found in many types of financial time series data where the assumption of conditional normality is violated. Finally, multivariate GARCH models are helpful tools for investigating volatility transmissions and patterns between two or more financial markets.

Although GARCH family models have time varying variance, the variance is not stochastic. Therefore, GARCH is not exactly the ARMA equivalent in the second moment. Stochastic volatility, as discussed in section 5.3, is not only time-varying, but also stochastic, and is probably the closest equivalent to an AR or ARMA process in the second moment.

5.1. ARCH and GARCH and their variations

5.1.1. ARCH and GARCH models

A stochastic process is called ARCH (AutoRegressive Conditional Heteroscedasticity) if its time varying conditional variance is heteroscedastic with autoregression:

$$y_t = \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2) \tag{5.1a}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 \tag{5.1b}$$

Equation (5.1a) is the mean equation where regressors can be generally added on to the right hand side alongside ε_t. Equation (5.1b) is the variance equation, which is an ARCH(q) process where autoregression in its squared residuals has an order of q, or has q lags.

A stochastic process is called GARCH (Generalised AutoRegressive Conditional Heteroscedasticity) if its time varying conditional variance is heteroscedastic with both autoregression and moving average:

$$y_t = \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2) \tag{5.2a}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_p^2$$

$$= \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{5.2b}$$

Equation (5.2) is a GARCH(p, q) process where autoregression in its squared residuals has an order of q, and the moving average component has an order of p.

One of the advantages of GARCH over ARCH is parsimonious, i.e. less lags are required to capture the property of time-varying variance in GARCH. In empirical applications a GARCH(1, 1) model is widely adopted. While in ARCH, for example, a lag length of five for daily data may still not be long enough. We demonstrate this with a GARCH(1, 1) model. Extending the variance process backwards yields:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

$$= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \left(\alpha_0 + \alpha_1 \varepsilon_{t-2}^2 + \beta_1 \sigma_{t-2}^2 \right)$$

$$= \cdots \cdots \tag{5.3}$$

$$= \frac{\alpha_0}{1 - \beta_1} + \alpha_1 \sum_{n=1}^{\infty} \beta_1^{n-1} \varepsilon_{t-n}^2$$

Indeed, only the first few terms would have noteworthy influence since $\beta_1^n \underset{n \to \infty}{\to} 0$. This shows how a higher-order ARCH specification can be approximated by a GARCH(1, 1) process.

Similar to ARMA models, there are conditions for stationarity to be met. As the name of the model suggests, the variances specified above are conditional. The unconditional variance of GARCH would be of interest to the property of the model. Applying the expectations operator to both sides of equation (5.2b), we have:

$$E(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} \alpha_i E(\varepsilon_{t-1}^2) + \sum_{j=1}^{p} \beta_j E(\sigma_{t-j}^2)$$

Noting $E(\sigma_t^2) = E(\varepsilon_{t-i}^2) = E(\sigma_{t-j}^2)$ is the unconditional variance of the residual, which is solved as:

$$\sigma^2 = E(\sigma_t^2) = \frac{\alpha_0}{1 - \sum\limits_{i=1}^{q} \alpha_i + \sum\limits_{j=1}^{p} \beta_j}$$

It is clear that for the process to possess a finite variance, the following condition must be met:

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1 \tag{5.4}$$

In commonly used GARCH(1, 1) models, the condition is simply $\alpha_1 + \beta_1 < 1$. Many financial time series have persistent volatility, i.e. the sum of α_i and β_j is close to being unity. A unity sum of α_i and β_j leads to so-called Integrated GARCH or IGARCH as the process is not covariance stationary. Nevertheless, this does not pose as serious a problem as it appears. According to Nelson (1990), Bougerol and Picard (1992) and Lumsdaine (1991), even if a GARCH (IGARCH) model is not covariance stationary, it is strictly stationary or ergodic, and the standard asymptotically based inference procedures are generally valid. See Chapter 1 of this book for various definitions of stationarity and ergodicity.

5.1.2. Variations of the ARCH/GARCH model

Variations are necessary to adapt the standard GARCH model to the need arising from examining the time series properties of specific issues in finance and economics. Here we present the model relating the return on a security to its time-varying volatility or risk – ARCH-M, and the models of asymmetry – Exponential GARCH (EGARCH) and Threshold GARCH (TGARCH).

The ARCH-M model

When the conditional variance enters the mean equation for an ARCH process, the ARCH-in-Mean or simply the ARCH-M model is derived:

$$y_t = \lambda_1 x_1 + \cdots \lambda_m x_m + \varphi \sigma_t^2 + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2) \tag{5.5a}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 \tag{5.5b}$$

where x_k, $k = 1, \ldots m$ are exogenous variables which could include lagged y_t. In the sense of asset pricing, if y_t is the return on an asset of a firm, then x_k, $k = 1, \ldots m$ would generally include the return on the market and possibly other

explanatory variables such as the price earnings ratio and the size. The parameter φ captures the sensitivity of the return to the time-varying volatility, or in other words, links the return to a *time-varying risk premium*. The ARCH-M model is generalised from the standard ARCH by Engle *et al.* (1987) and can be further generalised that the conditional variance is GARCH instead of ARCH, and that the conditional standard deviation, instead of the conditional variance, enters the mean equation.

The EGARCH model

The model captures asymmetric responses of the time-varying variance to shocks and, at the same time, ensures that the variance is always positive. It was developed by Nelson (1991) with the following specification:

$$\ln(\sigma_t^2) = \alpha_0 + \beta \ln(\sigma_{t-1}^2) + \alpha \left\{ \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| - \sqrt{\frac{2}{\pi}} \right\} - \gamma \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \tag{5.6}$$

where γ is asymmetric response parameter or leverage parameter. The sign of γ is expected to be positive in most empirical cases so that a negative shock increases future volatility or uncertainty while a positive shock eases the effect on future uncertainty. This is in contrast to the standard GARCH model where shocks of the same magnitude, positive or negative, have the same effect on future volatility. In macroeconomic analysis, financial markets and corporate finance, a negative shock usually implies bad news, leading to a more uncertain future. Consequently, for example, shareholders would require a higher expected return to compensate for bearing increased risk in their investment. A statistical asymmetry is, under various circumstances, also a reflection of the real world asymmetry, arising from the nature, process or organisation of economic and business activity, e.g. the change in financial leverage is asymmetric to shocks to the share price of a firm.

Equation (5.6) is, exactly speaking, an EGARCH(1, 1) model. Higher order EGARCH models can be specified in a similar way, e.g. EGARCH(p, q) is as follows:

$$\ln(\sigma_t^2) = \alpha_0 + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}^2) + \sum_{i=1}^{q} \left\{ \alpha_i \left(\left| \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right| - \sqrt{\frac{2}{\pi}} \right) - \gamma_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}} \right\} \tag{5.7}$$

The threshold GARCH model

It is also known as the GJR model, named after Glosten, Jagannathan and Runkle (1993). Despite the advantages EGARCH appears to enjoy, the empirical estimation of the model is technically difficult as it involves highly non-linear algorithms. In contrast, the GJR model is much simpler than, though not as

elegant as, EGARCH. A general GJR model is specified as follows:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \left\{ \alpha_i \varepsilon_{t-i}^2 + \delta_i \varepsilon_{t-i}^2 \right\} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{5.8}$$

where $\delta_i = 0$ if $\varepsilon_{t-i} > 0$. So, γ_i catches asymmetry in the response of volatility to shocks in a way that imposes a prior belief that for a positive shock and a negative shock of the same magnitude, future volatility is always higher, or at least the same, when the sign of the shock is negative. This may make sense under many circumstances but may not be universally valid. An alternative to the GJR specification is:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \left\{ \alpha_i^+ \varepsilon_{t-i}^2 + \alpha_i^- \varepsilon_{t-i}^2 \right\} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \tag{5.9}$$

where $\alpha_i^+ = 0$ if $\varepsilon_{t-i} < 0$, and $\alpha_i^- = 0$ if $\varepsilon_{t-i} > 0$. In such case, whether a positive shock or a negative shock of the same magnitude has larger effect on volatility will be subject to empirical examination.

5.2. Multivariate GARCH

We restrict our analysis to bivariate models as a multivariate GARCH with more than two variables would be extremely difficult to estimate technically and convey meaningful messages theoretically. A bivariate GARCH model expressed in matrices takes the form:

$$\mathbf{y}_t = \boldsymbol{\varepsilon}_t \tag{5.10a}$$

$$\boldsymbol{\varepsilon}_t \mid \boldsymbol{\Omega}_{t-1} \sim N(\mathbf{0}, \mathbf{H}_t) \tag{5.10b}$$

where vectors

$$\mathbf{y}_t = \begin{bmatrix} y_{1t} & y_{2t} \end{bmatrix}, \boldsymbol{\varepsilon}_t = \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{bmatrix}, \quad \text{and} \quad \mathbf{H}_t = \begin{bmatrix} h_{11t} & h_{12t} \\ h_{21t} & h_{22t} \end{bmatrix}$$

is the covariance matrix which can be designed in a number of ways. Commonly used specifications of the covariance include constant correlation, VECH (full parameterisation), and BEKK (positive definite parameterisation) named after Baba, Engle, Kraft and Kroner (1990). We introduce them in turn in the following.

5.2.1. *Constant correlation*

A constant correlation means that:

$$\frac{h_{12t}}{\sqrt{h_{11t} h_{22t}}} = \rho$$

is constant over time or it is not a function of time. Therefore, h_{12t} is decided as:

$$h_{12t} = \rho\sqrt{h_{11t}h_{22t}} \tag{5.11}$$

An obvious advantage in the constant correlation specification is simplicity. Nonetheless, it can only establish a link between the two uncertainties, failing to tell the directions of volatility spillovers between the two sources of uncertainty.

5.2.2. Full parameterisation

The full parameterisation, or VECH, converts the covariance matrix to a vector of variance and covariance. As $\sigma_{ij} = \sigma_{ji}$, the dimension of the vector converted from an $m \times m$ matrix is $m(m+1)/2$. Thus, in a bivariate GARCH process, the dimension of the variance/covariance vector is three. With a trivariate GARCH, the dimension of the vector is six, i.e. there are six equations to describe the time-varying variance/covariance. Therefore, it is unlikely to be feasible when more than two variables are involved in a system. The VECH specification is presented as:

$$\text{vech}(\mathbf{H}_t) = \text{vech}(\mathbf{A}_0) + \sum_{i=1}^{q} \mathbf{A}_i \text{vech}(\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}'_{t-i}) + \sum_{j=1}^{p} \mathbf{B}_j \text{vech}(\mathbf{H}_{t-j}) \tag{5.12}$$

where $\mathbf{H}_t$, $\mathbf{A}_0$, $\mathbf{A}_i$, $\mathbf{B}_j$ and $\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}'_t$ are matrices in their conventional form, and $\text{vech}(\cdot)$ means the procedure of conversion of a matrix into a vector, as described above. For $p = q = 1$, equation (5.12) can be written explicitly:

$$\begin{aligned}
\mathbf{H}_t = \begin{bmatrix} h_{11,t} \\ h_{12,t} \\ h_{22,t} \end{bmatrix} &= \begin{bmatrix} \alpha_{11,0} \\ \alpha_{12,0} \\ \alpha_{22,0} \end{bmatrix} + \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} & \alpha_{13,1} \\ \alpha_{21,1} & \alpha_{22,1} & \alpha_{23,1} \\ \alpha_{31,1} & \alpha_{32,1} & \alpha_{33,1} \end{bmatrix} \begin{bmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{bmatrix} \\
&+ \begin{bmatrix} \beta_{11,1} & \beta_{12,1} & \beta_{13,1} \\ \beta_{21,1} & \beta_{22,1} & \beta_{23,1} \\ \beta_{31,1} & \beta_{32,1} & \beta_{33,1} \end{bmatrix} \begin{bmatrix} h_{11,t-1} \\ h_{12,t-1} \\ h_{22,t-1} \end{bmatrix}
\end{aligned} \tag{5.13}$$

So, the simplest multivariate model has 21 parameters to estimate.

5.2.3. Positive definite parameterisation

It is also known as BEKK, suggested by Baba, Engle, Kraft and Kroner (1990). In fact, it is the most natural way to deal with multivariate matrix operations. The BEKK specification takes the following form:

$$\mathbf{H}_t = \mathbf{A}_0'\mathbf{A}_0 + \mathbf{A}_i'\,\boldsymbol{\varepsilon}_{t-i}\boldsymbol{\varepsilon}'_{t-i}\,\mathbf{A}_i + \mathbf{B}_j'\,\mathbf{H}_{t-j}\,\mathbf{B}_j \tag{5.14}$$

where $\mathbf{A}_0$ is a symmetric ($N \times N$) parameter matrix, and $\mathbf{A}_i$ and $\mathbf{B}_j$ are unrestricted ($N \times N$) parameter matrices. The important feature of this specification is that it

builds in sufficient generality, allowing the conditional variances and covariances of the time series to influence each other, and at the same time, does not require to estimate a large number of parameters. For $p = q = 1$ in a bivariate GARCH process, equation (5.14) has only 11 parameters compared with 21 parameters in the VECH representation. Even more importantly, the BEKK process guarantees that the covariance matrices are positive definite under very weak conditions; and it can be shown that under certain non-linear restrictions on $\mathbf{A}_i$ and $\mathbf{B}_j$, equation (5.14) and the VECH representation are equivalent (Engle and Kroner, 1995). In the bivariate system with $p = q = 1$, equation (5.14) becomes:

$$
\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} \alpha_{11,0} & \alpha_{12,0} \\ \alpha_{21,0} & \alpha_{22,0} \end{bmatrix}
$$

$$
+ \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix} \begin{bmatrix} \alpha_{11,1} & \alpha_{12,1} \\ \alpha_{21,1} & \alpha_{22,1} \end{bmatrix}
$$

$$
+ \begin{bmatrix} \beta_{11,1} & \beta_{12,1} \\ \beta_{21,1} & \beta_{22,1} \end{bmatrix}' \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{21,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} \beta_{11,1} & \beta_{12,1} \\ \beta_{21,1} & \beta_{22,1} \end{bmatrix} \quad (5.15)
$$

We can examine the sources of uncertainty and, moreover, assess the effect of signs of shocks with equation (5.15). Writing the variances and covariance explicitly:

$$
h_{11,t} = \alpha_{11,0} + (\alpha_{11,1}^2 \varepsilon_{1,t-1}^2 + 2\alpha_{11,1}\alpha_{21,1}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + \alpha_{21,1}^2 \varepsilon_{2,t-1}^2)
$$
$$
+ (\beta_{11,1}^2 h_{11,t-1} + 2\beta_{11,1}\beta_{21,1}h_{12,t-1} + \beta_{21,1}^2 h_{22,t-1}) \quad (5.16a)
$$

$$
h_{12,t} = h_{21,t} = \alpha_{12,0} + [\alpha_{11,1}\alpha_{12,1}\varepsilon_{1,t-1}^2 + (\alpha_{12,1}\alpha_{21,1} + \alpha_{11,1}\alpha_{22,1})\varepsilon_{1,t-1}\varepsilon_{2,t-1}
$$
$$
+ \alpha_{21,1}\alpha_{22,1}\varepsilon_{2,t-1}^2] + [\beta_{11,1}\beta_{21,1}h_{11,t-1} + (\beta_{12,1}\beta_{21,1}
$$
$$
+ \beta_{11,1}\beta_{22,1})h_{12,t-1} + \beta_{21,1}\beta_{22,1}h_{22,t-1}] \quad (5.16b)
$$

$$
h_{22,t} = \alpha_{22,0} + (\alpha_{12,1}^2 \varepsilon_{1,t-1}^2 + 2\alpha_{12,1}\alpha_{22,1}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + \alpha_{22,1}^2 \varepsilon_{2,t-1}^2)
$$
$$
+ (\beta_{12,1}^2 h_{11,t-1} + 2\beta_{12,1}\beta_{22,1}h_{12,t-1} + \beta_{22,1}^2 h_{22,t-1}) \quad (5.16c)
$$

Looking at the diagonal elements in the above matrix, i.e. $h_{11,t}$ and $h_{22,t}$, we can assess the impact of the shock in one series on the uncertainty or volatility of the other, and the impact could be asymmetric or only be one way effective. In particular, one might also be interested in assessing the effect of the signs of shocks in the two series. To this end the diagonal elements representing the previous shocks can be rearranged as follows:

$$
\alpha_{11,1}^2 \varepsilon_{1t-1}^2 + 2\alpha_{11,1}\alpha_{21,1}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + \alpha_{21,1}^2 \varepsilon_{2t-1}^2 = (\alpha_{11,1}\varepsilon_{1,t-1} + \alpha_{21,1}\varepsilon_{2,t-1})^2
$$
$$
(5.17a)
$$

$$
\alpha_{12,1}^2 \varepsilon_{1t-1}^2 + 2\alpha_{12,1}\alpha_{22,1}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + \alpha_{22,1}^2 \varepsilon_{2t-1}^2 = (\alpha_{12,1}\varepsilon_{1,t-1} + \alpha_{22,1}\varepsilon_{2,t-1})^2
$$
$$
(5.17b)
$$

It is clear that $\alpha_{11,1}$ and $\alpha_{22,1}$ represent the effect of the shock on the future uncertainty of the same time series and $\alpha_{21,1}$ and $\alpha_{12,1}$ represent the cross effect, i.e. the effect of the shock of the second series on the future uncertainty of the first series, and vice versa. The interesting point is that, if $\alpha_{11,1}$ and $\alpha_{21,1}$ have different signs, then the shocks with different signs in the two time series tend to increase the future uncertainty in the first time series. Similarly, if $\alpha_{12,1}$ and $\alpha_{22,1}$ have different signs, the future uncertainty of the second time series might increase if the two shocks have different signs. It seems that this model specification is appropriately fitted to investigate volatility spillovers between two financial markets.

The positive definite specification of the covariance extends the univariate GARCH model naturally, e.g. a BEKK-GARCH(1, 1) model can reduce to a GARCH(1, 1) when the dimension of the covariance matrix becomes one. Therefore, it is of interest to make inquiry into the conditions for covariance stationarity in the general matrix form. For this purpose, we need to vectorise the BEKK representation, i.e. to arrange the elements of each of the matrices into a vector. Due to the special and elegant design of the BEKK covariance, the vectorisation can be neatly and orderly derived, using one of the properties of vectorisation, i.e. vech$(ABC) = [C' \otimes A]$vech(B), where $\otimes$ is the Kroneker product. In this case, the innovation matrix;

$$\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1} = \begin{bmatrix} \varepsilon^2_{1,t-1} & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}\varepsilon_{1,t-1} & \varepsilon^2_2 \end{bmatrix}$$

and the covariance matrix:

$$\mathbf{H}_{t-1} = \begin{bmatrix} h_{1,t-1} & h^{1/2}_{1,t-1}h^{1/2}_{2,t-1} \\ h^{1/2}_{2,t-1}h^{1/2}_{1,t-1} & h_{2,t-1} \end{bmatrix}$$

are represented by B; and the fact that the parameter matrices $\mathbf{A}'$ and $\mathbf{A}$ and $\mathbf{B}'$ and $\mathbf{B}$ have already been transposed to each other further simplifies the transformation. For more details on these operations refer to Judge *et al.* (1988) and Engle and Kroner (1995). The vectorised $\mathbf{H}_t$ is derived as:

$$\text{vech}(\mathbf{H}_t) = (\mathbf{A}_0 \otimes \mathbf{A}_0)' \, \text{vech}(\mathbf{I}) + (\mathbf{A}_i \otimes \mathbf{A}_i)' \, \text{vech}\left(\boldsymbol{\varepsilon}_{t-1}\boldsymbol{\varepsilon}'_{t-1}\right)$$
$$+ \left(\mathbf{B}_j \otimes \mathbf{B}_j\right)' \, \text{vech}(\mathbf{H}_{t-1}) \tag{5.18}$$

the unconditional covariance is:

$$E(\mathbf{H}_t) = \left[\mathbf{I} - (\mathbf{A}_i \otimes \mathbf{A}_i)' - \left(\mathbf{B}_j \otimes \mathbf{B}_j\right)'\right]^{-} \text{vech}\left(\mathbf{A}'_0 \otimes \mathbf{A}_0\right) \tag{5.19}$$

and the conditions for covariance stationarity is:

$$\text{mod}\left[(\mathbf{A}_i \otimes \mathbf{A}_i)' + \left(\mathbf{B}_j \otimes \mathbf{B}_j\right)'\right] < 1 \tag{5.20}$$

That is, for $\boldsymbol{\varepsilon}_t$ to be covariance stationary, all the eigenvalues of $(\mathbf{A}_i \otimes \mathbf{A}_i)' + \left(\mathbf{B}_j \otimes \mathbf{B}_j\right)'$ are required to be less than one in modules. There are altogether four

eigenvalues for a bivariate GARCH process as the Kroneker product of two (2×2) matrices produces a (4×4) matrix. These eigenvalues would be complex numbers in general. When the dimension of the covariance is one, equation (5.20) reduces to equation (5.4) for the univariate case.[1]

5.3. Stochastic volatility

ARCH/GARCH processes are not really stochastic, rather they are deterministic and the conditional variance possesses no unknown innovations at the time. ARCH and GARCH are not exactly the second moment equivalent to AR and ARMA processes in the mean. Stochastic volatility, as favoured by Harvey *et al.* (1994), Ruiz (1994), Andersen and Lund (1997) and others, is probably the closest equivalent to an AR or ARMA process in describing the dynamics of variance/covariance. Let us look at a simple case:

$$y_t = \sigma_t \varepsilon_t$$
$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{5.21}$$
$$h_t = \ln \sigma_t^2 \sim \text{ARMA}(q, p)$$

The logarithm of the variance in a stochastic volatility model, $h_t = \ln \sigma_t^2$, behaves exactly as a stochastic process in the mean, such as random walks or AR or ARMA processes. For example, if h_t is modelled as an AR(1) process, then:

$$h_t = \alpha + \rho h_{t-1} + v_t$$
$$v_t \sim N(0, \sigma_v^2) \tag{5.22}$$

Alternatively when h_t is modelled as an ARMA(1, 1) process:

$$h_t = \alpha + \rho h_{t-1} + v_t + \theta v_{t-1}$$
$$v_t \sim N(0, \sigma_v^2) \tag{5.23}$$

When the stochastic part of volatility, v_t, does not exist (i.e. $\sigma_v^2 = 0$), equation (5.22) does not reduce to ARCH(1) but to GARCH(1, 0). So the difference in modelling variance is substantial between GARCH and stochastic volatility approaches. To estimate stochastic volatility models, expressing equation (5.21) as:

$$g_t = h_t + \kappa_t \tag{5.24}$$

where $g_t = \ln(y_t^2)$, and $\kappa_t = \ln(\varepsilon_t^2)$. We can see that h_t becomes part, or a component, of the (transformed) time series, in contrast to traditional statistical models where the variance expresses the distribution of variables in a different way. As the time series now has more than one component, neither is readily observable, so the components are often referred to as unobserved components. These components together form the whole system and individually describe the

state of the system from certain perspectives, so they can be referred to as state variables as well. Such a specification poses problems as well as advantages: decomposition into components can be arbitrary and estimation can be complicated and sometimes difficult; nevertheless, the state variables and their dynamic evolution and interaction may reveal the fundamental characteristics and dynamics of the system or the original time series more effectively, or provide more insights into the working of the system. Models of this type are usually estimated in the state space, often accompanied by the use of Kalman filters. See Chapter 9 for details of the state space representation and the Kalman filter.

5.4. Examples and cases

When time comes up to implementing an empirical study, the problem may never be exactly the same as illustrated in the text. This is hopefully what a researcher expects to encounter rather than attempting to avoid if s/he imagines new discoveries in her/his study or would like to differentiate her/his study from others. This section provides such examples.

Example 5.1

This is an example incorporating macroeconomic variables into the conditional variance equation for stock returns by Hasan and Francis (1998), entitled 'Macroeconomic factors and the asymmetric predictability of conditional variances'. The paper includes the default premium, dividend yield and the term premium as state variables in the conditional variance equation, though its main purpose is to investigate the predictability of the volatilities of large versus small firms. The paper shows that volatility surprises of small (large) firms are important in predicting the conditional variance of large (small) firms, and this predictive ability is still present when the equation of conditional variance includes above mentioned state variables.

   The paper uses monthly returns of all NYSE and AMEX common stocks with year-end market value information available gathered from the Center for Research in Security Prices (CRSP) monthly master tape from 1926 to 1988. All stocks in the sample are equally divided into twenty size-based portfolios, S1 (smallest) to S20 (largest), according to the market value of equity at the end of the prior year. Monthly excess returns on each of the portfolios are obtained by averaging returns across all stocks included in the portfolio. Their specification is as follows:

$$R_{i,t} = \alpha_{i,t} + \beta_i R_{i,t-1} + \mu_{i,1} JAN_t + \gamma_i R_{j,t-1} + e_{i,t} \qquad (5.25a)$$

$$h_{i,t} = \delta_{i,0} + \alpha_i e_{i,t-1}^2 + \theta_i h_{i,t-1} + \delta_{i1} JAN_t + \varphi_j e_{j,t-1}^2 + \sum \omega_k Z_{k,t-1}$$
$$\qquad (5.25b)$$

Continued

The mean equation follows an AR(1) process. JAN_t is the dummy which is equal to one when in January and zero otherwise. $Z_{k,t}(k = 1, 2, 3)$ are the state variables of default premium (DEF), dividend yield (DYLD) and the term premium (TERM) respectively. These state variables are those used by Fama and French (1989) and Chen (1991). The effect of return and volatility spillovers across portfolios is through the inclusion of lagged returns on portfolio j in the mean equation for portfolio i, and the inclusion of lagged squared errors for portfolio j in the conditional variance equation of portfolio i. Squared errors for portfolio j are obtained through estimating a basic GARCH model whose conditional variance is a standard GARCH(1, 1) plus the January dummy. Therefore, the model is univariate rather than bivariate in nature.

The paper then estimates the model for two portfolios, the small size stock portfolio (Table 5.1) and large size stock portfolio (Table 5.2). Volatility spillovers across these two portfolios are examined. The major findings are that while return spillovers are from the small stock portfolio to the large stock portfolio only, volatility spillovers are bi-directional, though the effect of the small stock portfolio on the large stock portfolio is greater than that of the other way round. Only the main part of the results is presented in the following tables. Model (1) does not include the state variables,

Table 5.1 Small stock portfolio

Mean equation

μ_0	$R_{1,t-1}$	JAN	$R_{20,t-1}$
0.0048	0.1896	0.1231	0.0287
(1.705)	(4.679)	(8.163)	(0.475)

Variance equation

Model	δ_0	$e_{1,t-1}^2$	$h_{1,t-1}$	JAN	$e_{20,t-1}^2$	DYLD	TERM	DEF
(1)	−0.0001	0.0284	0.9305	0.0005	0.0022			
	(0.998)	(2.039)	(41.915)	(0.464)	(2.741)			
(2)	−0.0001	0.0259	0.9341	0.0008	0.0021	−0.003		
	(0.470)	(1.890)	(43.648)	(0.738)	(2.629)	(1.167)		
(3)	−0.0001	0.0254	0.9316	0.0009	0.0023		−0.0001	
	(0.228)	(1.812)	(43.310)	(0.789)	(2.650)		(2.991)	
(4)	−0.0001	0.0273	0.9286	0.0008	0.0024			−0.0001
	(0.134)	(1.843)	(39.142)	(0.729)	(2.681)			(3.014)
(5)	−0.0001	0.0009	0.9694	0.0012	0.0018	−0.0005	−0.0001	−0.0001
	(0.740)	(0.092)	(82.117)	(1.193)	(3.556)	(4.160)	(0.693)	(0.751)

Robust *t*-statistics in parentheses.

Table 5.2 Large stock portfolio

Mean equation

μ_0	$R_{20,t-1}$	JAN	$R_{1,t-1}$
0.0064	0.0497	−0.0082	0.1013
(3.867)	(1.295)	(1.413)	(6.051)

Variance equation

Model	δ_0	$e_{20,t-1}^2$	$h_{20,t-1}$	JAN	$e_{1,t-1}^2$	DYLD	TERM	DEF
(1)	0.0002	0.1610	0.6511	−0.0001	0.0237			
	(2.981)	(3.347)	(10.148)	(0.194)	(2.986)			
(2)	0.0003	0.1595	0.6504	−0.0001	0.0239	−0.0001		
	(2.563)	(3.324)	(9.926)	(0.226)	(3.043)	(0.795)		
(3)	0.0001	0.1544	0.6293	−0.0001	0.0282		0.0004	
	(1.811)	(3.252)	(9.986)	(0.205)	(3.442)		(2.492)	
(4)	0.0001	0.1549	0.6460	−0.0003	0.0244			0.0003
	(1.321)	(3.214)	(9.567)	(0.089)	(4.489)			(1.989)
(5)	0.0001	0.1502	0.6322	−0.0001	0.2974	0.0002	0.0008	−0.0001
	(0.820)	(3.270)	(10.214)	(0.393)	(3.733)	(0.784)	(2.055)	(1.111)

Robust *t*-statistics in parentheses.

Models (2)–(4) include one of the state variables each, and Model (5) incorporates all the state variables. As there is not much difference in the mean equation results, only the results from Model (5) are provided.

Example 5.2

This is an example of the bivariate GARCH model applied to the foreign exchange market by Wang and Wang (2001). In this study, the daily spot and forward foreign exchange rates of the British pound, German mark, French franc and Canadian dollar against the US dollar are used. All of the data sets start from 02/01/76 and end on 31/12/90; so there are 3,758 observations in each series. These long period high frequency time series data enable us to observe a very evident GARCH phenomenon in a bivariate system. The system of equations for the spot exchange rate, S_t, and the forward exchange rate, F_t, is specified as an extended VAR, which incorporates a forward premium into a simple VAR. In addition, the covariance of the extended VAR is time-varying which allows for and mimics volatility spillovers

Continued

or transmission between the spot and forward foreign exchange markets. The model is given as follows:

$$\Delta s_t = c_1 + \gamma_1(f_{t-1} - s_{t-1}) + \sum_{i=1}^{m} \alpha_{1i} \Delta s_{t-i} + \sum_{i=1}^{m} \beta_{1i} \Delta f_{t-i} + \varepsilon_{1t}$$

$$\Delta f_t = c_2 + \gamma_2(f_{t-1} - s_{t-1}) + \sum_{i=1}^{m} \alpha_{2i} \Delta s_{t-i} + \sum_{i=1}^{m} \beta_{2i} \Delta f_{t-i} + \varepsilon_{2t}$$

(5.26)

$$\varepsilon_t \mid \Omega_{t-1} \sim N(0, \ H_t)$$

where $s_t = \mathrm{Ln}(S_t), f_t = \mathrm{Ln}(F_t)$, $\Delta s_t = s_t - s_{t-1}$, $\Delta f_t = f_t - f_{t-1}$, and H_t is the time-varying covariance matrix with the BEKK specification.

The inclusion of the forward premium is not merely for setting up an ECM model, it keeps information in levels while still meeting the requirements for stationarity. Although there are arguments about the property of the forward premium, its inclusion makes the system informationally and economically complete by reserving information in levels (original variables) and reflecting expectations in the market.

The bivariate GARCH effects are, in general, strong in both the spot and forward markets, though there exists a clear asymmetry in the volatility spillover patterns. That is, there are volatility spillovers from the spot market to the forward, to a lesser extent, compared with the other way round. Table 5.3 presents the results based mainly on the second moment.

In addition, the parameter for the forward premium is also reported, as it would validate the cointegration between the spot and forward exchange rates and the need to incorporate the forward premium. Consider the British pound first. a_{12} and a_{21} are both significant at 1 per cent level, but the magnitude of the former is about half the size of the latter, implying that the effect of the shock in the forward market on the spot market volatility is bigger than that on the forward market induced by the shock in the spot market. Turning to the effects of the previous uncertainty, while b_{21} is significant, b_{12} is not significant at all, so the volatility spillovers are one directional from the forward to the spot. Notice, b_{22} is also insignificant, which means there is only ARCH in the forward exchange rate. Further scrutiny on the signs of a_{12} and a_{22} suggests that the future volatility in the forward market would be higher if the two shocks have different signs. In the case of the German mark, the asymmetry is more apparent, where a_{12} is not significant at all but a_{21} is significant at 1 per cent level. As such, the shock in the forward market would affect the future volatility in the spot market, but the shock in the spot market has no influence on the future volatility in the forward market. In addition, a_{11} and a_{21} have different signs, so the

Table 5.3 Volatility spillovers between spot and forward FX rates

$$\Delta s_t = c_1 + \gamma_1(f_{t-1} - s_{t-1}) + \sum_{i=1}^{m} \alpha_{1i}\Delta s_{t-i} + \sum_{i=1}^{m} \beta_{1i}\Delta f_{t-i} + \varepsilon_{1t}$$

$$\varepsilon_t \,|\, \Omega_{t-1} \sim N(0, H_t)$$

$$\Delta f_t = c_2 + \gamma_2(f_{t-1} - s_{t-1}) + \sum_{i=1}^{m} \alpha_{2i}\Delta s_{t-i} + \sum_{i=1}^{m} \beta_{2i}\Delta f_{t-i} + \varepsilon_{2t}$$

$$\begin{bmatrix} h_{11,t} & h_{12,t} \\ h_{21,t} & h_{22,t} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}' \begin{bmatrix} \varepsilon_{1,t-1}^2 & \varepsilon_{1,t-1}\varepsilon_{2,t-1} \\ \varepsilon_{1,t-1}\varepsilon_{2,t-1} & \varepsilon_{2,t-1}^2 \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}' \begin{bmatrix} h_{11,t-1} & h_{12,t-1} \\ h_{12,t-1} & h_{22,t-1} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

	BP	DM	FF	CD
c_1	0.00025**	−0.00058***	0.00012	0.00013**
	(2.3740)	(4.3591)	(1.2106)	(2.2491)
γ_1	−0.12034***	−0.23731***	−0.03214	−0.05255
	(3.9301)	(5.5473)	(1.3884)	(1.8081)*
c_2	0.00027**	−0.00057***	0.00014	0.00015**
	(2.5077)	(4.0781)	(1.4735)	(2.5582)
γ_2	−0.12536***	−0.23203***	−0.05197**	−0.05436*
	(4.0293)	(5.1272)	(2.2460)	(1.8156)
a_{11}	0.51775***	−0.20555***	1.00020***	0.53282***
	(13.3029)	(3.8590)	(148.5928)	(8.4157)
a_{12}	−0.24576***	0.00539	0.03776***	−0.05055
	(6.8053)	(0.1062)	(5.4583)	(0.8144)
a_{21}	0.45452***	1.17328***	−0.06149***	0.40725***
	(11.5872)	(22.2677)	(8.7868)	(6.5068)
a_{22}	1.21688***	0.96138***	0.89990***	0.98706***
	(33.3811)	(19.1147)	(124.5143)	(16.0892)
b_{11}	0.43475***	1.00966***	0.24888***	0.52226***
	(4.8987)	(16.8606)	(8.0209)	(8.0757)
b_{12}	−0.10742	−0.18033***	−0.04565	0.01389
	(1.2389)	(3.5195)	(1.4180)	(0.2278)
b_{21}	−0.66683***	−1.16582***	0.10881***	−0.22376***
	(7.5985)	(19.1740)	(3.5734)	(3.5935)
b_{22}	−0.13151	−0.05456	0.40644***	0.28611***
	(1.5358)	(1.0563)	(12.7895)	(4.7730)

t-statistics in parentheses.
*significant at 10 per cent level; **significant at 5 per cent level; ***significant at 1 per cent level.
Constant terms in the second moment are not reported.

Continued

shock with opposite signs in these two markets would be inclined to increase
the future volatility in the spot market. As far as the previous variance is
concerned, b_{12} and b_{21} are both significant, but the size of the former is
much smaller than that of the latter, so the asymmetry exists in this respect
too. Again, b_{22} is not significant; the forward rate would only have the
ARCH effect if it were not considered in a bivariate system. The strongest
asymmetry occurs in the exchange rates of the Canadian dollar. The volatility
spillovers are absolutely one directional from the forward rate to the spot
rate. That is: a_{12} and b_{12} are not significant at any conventional levels,
whereas a_{21} and b_{21} are both significant at 1 per cent level. Similar to
the British pound, in the case of the French franc, the influence of the
previous variance is clearly one directional from the forward to the spot
measured by b_{12} and b_{21}. Although the GARCH effect is strong in the
forward rate as well as in the spot rate, b_{22} is close to being twice as big
as b_{11}. Regarding the previous shocks, the influence is also more from
forward to spot; both a_{12} and a_{21} are significant but a_{21} is much bigger
than a_{12}. Therefore, the four currencies have similar asymmetric volatility
spillover patterns. Another interesting point in the franc example is that the
premium is not significant in both the spot and forward equations when the
covariance matrix is assumed as constant. The premium is significant in
the forward equation when estimated in a multivariate GARCH framework.
This suggests that the rejection/acceptance of a cointegration relationship
is, to a certain extent, subject to the assumption on the properties of the
covariance.

In Table 5.4, all four eigenvalues for each currency are reported. Their
positioning on the complex plane is displayed in Figure 5.1. It can been
seen that the biggest of the eigenvalues for each currency is around 0.96
in modules, so the time varying volatility is highly persistent. In the French

Table 5.4 Verifying covariance stationarity: the eigenvalues. Unconditional
covariance: $E(\sigma_t^2) = [I - (A^* \otimes A^*)' - (B^* \otimes B^*)']^{-1} \operatorname{vec}(C_0^{*'} C_0^*)$

$(A^* \otimes A^*)' + (B^* \otimes B^*)'$	BP	DM	FF	CD
λ_1 (real, imaginary)	0.963, 0.000	0.969, 0.000	1.003, 0.000	0.969, 0.000
λ_1 (mod)	0.963	0.969	1.003	0.969
λ_2 (real, imaginary)	0.852, 0.000	0.570, 0.000	0.995, -0.022	0.699, 0.000
λ_2 (mod)	0.852	0.570	0.996	0.699
λ_3 (real, imaginary)	0.628, 0.000	0.022, 0.000	0.995, 0.022	0.698, 0.000
λ_3 (mod)	0.628	0.022	0.996	0.698
λ_4 (real, imaginary)	0.608, 0.000	0.017, 0.000	0.988, 0.000	0.600, 0.000
λ_4 (mod)	0.608	0.017	0.988	0.600

In a situation that all eigenvalues are smaller than one in modules, the covariance is confirmed
stationary.

Eigenvalues of covariance matrices on the complex plane
(the horizontal axis is for the real part, and the vertical axis is for the imaginary
part of the eigenvalue. the reference circle is the unit circle)

Figure 5.1 Eigenvalues on the complex plane.

franc case, the biggest module of eigenvalue is just above unity, suggesting that the unconditional covariance does not exist. There are two explanations to provide for this. First, according to Nelson (1990), Bougerol and Picard (1992) and Lumsdaine (1991), even if a GARCH (IGARCH) model is not covariance stationary, it is strictly stationary or ergodic, and the standard asymptotically based inference procedures are generally valid. Second, the derivation of eigenvalues is based on the assumption that the spot variance and forward variance are equal in size. Nevertheless, the forward variance is smaller than the spot variance in the French franc case. Taking this into account, all of the modules of eigenvalue for the French franc become

Continued

less than one and the covariance stationary exists. The analysis on the eigenvalues of the Kroneker product of the covariance matrices reveals that the time varying volatility is also highly persistent in a bivariate setting for foreign exchange rate data. In addition, though the BEKK specification has proved a helpful analytical technique for volatility transmissions, especially the impact of the signs of the shocks in different markets, in empirical research, the covariance stationarity is not so easy to satisfy and is not always guaranteed.

5.5. Empirical literature

While time-varying volatility has found applications in almost all time series modelling in economics and finance, it attracts most attention in the areas of financial markets and investment where vast empirical literature has been generated, which has in turn brought about new forms and variations of this family of models. Time-varying volatility has become the norm in financial time series modelling, popularly accepted and applied by academics and professionals alike since the 1990s. Moreover, analysis of interactions between two or more variables in the first moment, such as in VAR and ECM, is extended through the use of time-varying volatility models, to the second moment to examine such important issues as volatility spillovers or transmissions between different markets.

   One of the most extensively researched topics is time-varying volatility universally found in stock market indices. Although findings vary from one market to another, a pattern of time-varying volatility, which is also highly persistent, is common to most of them. Nevertheless, many of studies attempt to exploit new features and add variations in model specifications to meet the specific need of empirical investigations. To examine the characteristics of market opening news, Gallo and Pacini (1998) apply the GARCH model and evaluate the impact of the news on the estimated coefficients of the model. They find that the differences between the opening price of one day and the closing price of the day before have different characteristics and have the effect of modifying the direct impact of daily innovations on volatility which reduces the estimated overall persistence of innovations. It is also claimed that the inclusion of this news variable significantly improves out-of-sample forecasting, compared with the simple GARCH model's performance. Brooks *et al.* (2000) adopt the power ARCH (PARCH) model proposed by Ding *et al.* (1993) to stock market returns in ten countries and a world index. As PARCH removes the restriction implicitly imposed by ARCH/GARCH, i.e. the power transformation is achieved by taking squaring operations of the residual or to the power of 2, it can possess richer volatility patters such as asymmetry and leverage effects. They find that the PARCH model is applicable to these return indices and that the optimal power transformation is remarkably similar across countries. Longin (1997) employs the analytical framework of

Kyle (1985) where there are three types of traders: informed traders, liquidity traders and market makers. In such a setting, the paper models information as an asymmetric GARCH process that large shocks are less persistent in volatility than small shocks. This, it is claimed, allows one to derive implications for trading volume and market liquidity. The study by Koutmos (1992) is one of the typical empirical applications of GARCH in finance in early times – risk–return trade-off in a time-varying volatility context and asymmetry of the conditional variance in response to innovations. The Exponential GARCH in Mean (EGARCH-M) model is chosen for above obvious reasons and the findings support the presence of these well observed phenomena in ten stock market return indices. Newly added to this literature is evidence from so-called emerging markets and the developing world. Investigating the behaviour of the Egyptian stock market in the context of pricing efficiency and the return-volatility relationship, Mecagni and Sourial (1999) employ a GARCH-M model to estimate four daily indices. Their results suggest that there is tendency of volatility clustering in returns, and a positive but asymmetric link between risk and returns which is statistically significant during market downturns. They claim that the asymmetry in the risk–return relationship is due to the introduction of circuit breakers. Husain (1998) examines the Ramadhan effect in the Pakistani stock market using GARCH models. Ramadhan, the season of the holy month of fasting, is expected to have effects on stock market behaviour one way or another. The study finds that the market is indeed tranquil as the conditional variance declines in that month, but the season does not appear to have impact on mean returns. Applying TGARCH models to two Eastern European markets, Shields (1997) reports findings contrary to those in the west that there is no asymmetry in the conditional variance in response to positive and negative shocks in these Eastern European markets.

International stock market linkages have attracted increasing attention in the process of so-called globalisation in a time when there are no major wars. Seeking excess returns through international diversification is one of the strategies employed by large multinational financial institutions in an ever intensifying competitive financial environment, while national markets, considered individually, appear to have been exploited to their full so that any non-trivial profitable opportunities do not remain in the context of semi-strong market efficiency. In particular, US investors have gradually given up the stand of regarding foreign markets as alien lands and changed their risk perspectives – international diversification benefits are more than off-setting perceived additional risks. In the meantime, international asset pricing theory has been developed largely with a stratified approach which regards the international financial market as segmented as well as linked markets, adding additional dimensions to the original capital asset pricing model which is, ironically, universal, or in other words, global. Under such circumstances, it is not strange that applications of multivariate GARCH models have mushroomed during this period.

Investigating one of the typical features in emerging financial markets, Fong and Cheng (2000) test the information based hypothesis that the rate of information absorption in the conditional variance is faster for foreign shares (open to foreigners

and locals) than for local shares (open to locals only) using bivariate GARCH(1, 1) model for nine dual listed stocks over the period 1991–1996. Their evidence indicates that the rate of information absorption is consistent with what proposed by Longin (1997), that the rate of information absorption varies inversely with the number of informed traders. They claim that removing foreign ownership restrictions is likely to improve both market efficiency and liquidity. International risk transmission or volatility spillovers between two or more financial markets is by far the most intensively researched area. In this fashion, Kim and Rui (1999) examine the dynamic relationship among the US, Japan and UK daily stock market return volatility and trading volume using bivariate GARCH models. They find extensive and reciprocal volatility spillovers in these markets. The results from return spillovers, or Granger causality in the mean equations, seem to confirm all reciprocal relationship but exclude London's influence on the New York Stock Exchange. Tay and Zhu (2000) also find such dynamic relationship in returns and volatilities in Pacific-Rim stock markets. Chou *et al.* (1999) test the hypothesis that the short-term volatility and price changes spill over from developed markets to emerging markets using the US and Taiwan data. They find substantial volatility spillover effect from the US stock market to the Taiwan stock market, especially for the model using close-to-open returns. There is also, it is claimed, evidence supporting the existence of spillovers in price changes. In contrast to the majority of the findings, Niarchos *et al.* (1999) show that there are no spillovers in means and conditional variances between the US and Greek stock markets and suggest that the U.S. market does not have a strong influence on the Greek stock market. Many similar studies have emerged in recent years, for example, Dunne (1999) and Darbar and Deb (1997), to mention a few.

  Inflation uncertainty remains one of the main application areas of GARCH modelling, following the first paper of this type on the topic by Engle (1982). In a recent study, Grier and Perry (1998), without much surprise, provide empirical evidence that inflation raises inflation uncertainty, as measured by the conditional variance of the inflation rate, for all G7 countries in the period from 1948 to 1993. Their results on the causal relationship from inflation uncertainty to inflation are mixed. In three countries, increased inflation uncertainty lowers inflation; while in two countries increased inflation uncertainty raises inflation. These findings have been extended to cover the developing world as well. Applying a similar testing procedure, Nas and Perry (2000) find evidence supporting the claim that inflation raises inflation uncertainty in Turkey over the full sample period of 1960 to 1998 and in the three sub-samples. They again show mixed results for the effect of inflation uncertainty on inflation, and claim that this is due to institutional and political factors in the monetary policy making process in Turkey between 1960 and 1998. Wang *et al.* (1999) examine the causal relationships between inflation, inflation uncertainty as measured with the conditional variance of the aggregate inflation rate, and relative price variability in sectoral price indices. They find that, although inflation does Granger cause inflation uncertainty, relative price variability is more a source of inflation uncertainty than the inflation level

itself. In contrast, Grier and Perry (1996) present different findings in respect of these relationships and appear to contradict the results of their other studies. Various studies on the topic include Brunner and Hess (1993), and Loy and Weaver (1998).

On foreign exchange markets, time-varying volatility models have been widely adopted to study various issues ranging from time-varying risk premia, volatility spillovers between the spot and forward exchange market, hedging strategies, to the effect of monetary policy. Searching for an explanation for the departure from Uncovered Interest Parity (UIP), Tai (1999) examines the validity of the risk premium hypothesis using a GARCH-M(1, 1) model. The empirical evidence supports the notion of time-varying risk premia in explaining the deviations from UIP. It also supports the idea that the foreign exchange risk is not diversifiable and hence should be priced in both foreign exchange market and equity market. Hu's (1997) approach is to examine the influence of macroeconomic variables on foreign exchange risk premia. The paper assumes that money and production follow a joint stochastic process with bivariate GARCH innovations based on Lucas's asset pricing model and implies that the risk premium in the foreign exchange market is due to time-varying volatilities in macroeconomic variables. Testing the model for three currencies shows that the time-varying risk premium is able to explain the deviation of the forward foreign exchange rate from the future spot rate. It is claimed that the model partially supports the efficient market hypothesis after accounting for time-varying risk premia. Investigating the effect of central bank intervention, Dominguez (1993) adopts GARCH models to test whether the conditional variance of exchange rates has been influenced by the intervention. The results indicate that intervention need not be publicly known for it to influence the conditional variance of exchange rate changes. Publicly known Fed intervention generally decreases exchange rate volatility, while secret intervention operations by both the Fed and the Bundesbank generally increase the volatility. Kim and Tsurumi (2000), Wang and Wang (1999), Hassapis (1995), Bollerslev and Melvin (1994), Copeland and Wang (1993), Mundaca (1991), Bollerslev (1990) and many other studies are also in this important area.

As mentioned earlier time-varying volatility has become the norm in financial time series modelling, popularly accepted and applied by academics and professionals alike since the 1990s. Therefore it does not appear to be feasible to exhaust listing the application areas and individual cases. Among other things not covered by the brief survey in this section, there are applications in option modelling, dynamic hedging, the term structure, interest rates and interest rate related financial instruments.

Note

1 Equation (5.14) becomes $h_{11,t} = \alpha_{11,0}^2 + \sum_{i=i}^{q} \alpha_{11,i}^2 \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_{11,j}^2$, so $\alpha_{11,0}^2$, $\alpha_{11,i}^2$ and $\beta_{11,j}^2$ are equivalent to α_0, α_i and β_j in equation (5.4) respectively.

Questions and problems

1   Describe ARCH and GARCH in comparison with AR and ARMA in the mean process.
2   Discuss many variations of GARCH and their relevance to financial modelling.
3   What is the stochastic volatility model? Discuss the similarities and differences between a GARCH type model and a stochastic volatility model.
4   Compare different specifications of multivariate GARCH models and comment on their advantages and disadvantages.
5   Collect data from Datastream to test for GARCH phenomena, using the following time series:

   (a) foreign exchange rates of selected industrialised nations and developing economies *vis-à-vis* the US$, taking the log or log difference transformation if necessary prior to the test,
   (b) CPI of the UK, US, Japan, China, Russia, and Brazil, taking any necessary transformation prior to the test,
   (c) total return series of IBM, Microsoft, Sage, Motorola, Intel, Vodafone, and Telefonica, taking any necessary transformation prior to the test.

   What do you find of their characteristics?
6   Collect data from Datastream and apply various multivariate GARCH models to the following time series:

   (a) the spot and forward foreign exchange rates of selected industrialised nations and developing economies *vis-à-vis* the US$, taking the log or log difference transformation if necessary prior to the test,
   (b) the stock market return indices of the US (e.g. S&P500) and the UK (e.g. FTSE100),
   (c) the stock market return indices of Japan and Hong Kong.

   What do you find of their links in the second moment?
7   Discuss and comment on the new developments in modelling time-varying volatilities.

References

Andersen, T.G. and Lund, J. (1997), Estimating continuous-time stochastic volatility models of the short-term interest rate, *Journal of Econometrics*, **77**, 343–377.

Baba, Y., Engle, R.F., Kraft, D.F. and Kroner, K.F. (1990), Multivariate simultaneous generalised ARCH, *Mimeo*, Department of Economics, University of California, San Diego.

Bollerslev, T. (1986), Generalized autoregressive conditional heteroskedasticity, *Journal of Econometrics*, **31**, 307–327.

Bollerslev, T. (1990), Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model, *Review of Economics and Statistics*, **72**, 498–505.

Bollerslev, T. and Melvin, M. (1994), Bid-ask spreads and volatility in the foreign exchange market: an empirical analysis, *Journal of International Economics*, **36**, 355–372.

Bougerol, P. and Picard, N. (1992), Stationarity of GARCH processes and some nonnegative time series, *Journal of Econometrics*, **52**, 115–127.

Brooks, R.D., Faff, R.W., McKenzie, M.D. and Mitchell, H. (2000), A multi-country study of power arch models and national stock market returns, *Journal of International Money and Finance*, **19**, 377–397.

Brunner, A.D. and Hess, G.D. (1993), Are higher levels of inflation less predictable? A state-dependent conditional heteroscedasticity approach, *Journal of Business and Economic Statistics*, **11**, 187–197.

Chen, N.F. (1991), Financial investment opportunities and the macroeconomy, *Journal of Finance*, **46**, 529–554.

Chou, R.Y., Lin, J.L. and Wu, C.S. (1999), Modeling the Taiwan stock market and international linkages, *Pacific Economic Review*, **4**, 305–320.

Copeland, L. and Wang, P.J. (1993), Estimating daily seasonals in financial time series: the use of high-pass spectral filters, *Economics Letters*, **43**, 1–4.

Darbar, S.M. and Deb, P. (1997), Co-movements in international equity markets, *Journal of Financial Research*, **20**, 305–322.

Ding, Z., Granger, C.W.J. and Engle, R.F. (1993), A long memory property of stock market returns and a new model, *Journal of Empirical Finance*, **1**, 83–106.

Dominguez, K.M. (1993), Does central bank intervention increase the volatility of foreign exchange rates? *National Bureau of Economic Research Working Paper*: 4532.

Dunne, P.G. (1999), Size and book-to-market factors in a multivariate GARCH-in-Mean asset pricing application, *International Review of Financial Analysis*, **8**, 35–52.

Engle, R.F. (1982), Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, **50**, 987–1007.

Engle, R.F. and Kroner, K.F. (1995), Multivariate simultaneous generalized ARCH, *Econometric Review*, **11**, 122–150.

Engle, R.F., Lilien, D.M. and Robins, R.P. (1987), Estimating time varying risk premia in the term structure: the ARCH-M model, *Econometrica* **55**, 391–407.

Fama, E. and French, K. (1989), Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics*, **25**, 23–49.

Fong, W.M. and Cheng, P.L. (2000), On the rate of information absorption in the conditional variance of SES dual listed stocks, *International Journal of Theoretical and Applied Finance*, **3**, 205–217.

Gallo, G.M. and Pacini, B. (1998), Early news is good news: the effects of market opening on market volatility, *Studies in Nonlinear Dynamics and Econometrics*, **2**, 115–131.

Glosten, L.R., Jagannathan, R. and Runkle, D. (1993), On the relation between the expected value and the volatility of the normal excess return on stocks, *Journal of Finance*, **48**, 1779–1801.

Grier, K.B. and Perry, M.J. (1996), Inflation, inflation uncertainty, and relative price dispersion: evidence from bivariate GARCH-M models, *Journal of Monetary Economics*, **38**, 391–405.

Grier, K.B. and Perry, M.J. (1998), On inflation and inflation uncertainty in the G7 countries, *Journal of International Money and Finance*, **17**, 671–689.

Harvey, A., Ruiz, E. and Shephard, N. (1994), Multivariate stochastic variance models, *Review of Economic Studies*, **61**, 247–264.

Hasan, I. and Francis, B.B. (1998), Macroeconomic factors and the asymmetric predictability of conditional variances, *European Financial Management*, **4**, 207–230.

Hassapis, C. (1995), Exchange risk in the EMS: some evidence based on a GARCH model, *Bulletin of Economic Research*, **47**, 295–303.

Hu, X.Q. (1997), Macroeconomic uncertainty and the risk premium in the foreign exchange market, *Journal of International Money and Finance*, **16**, 699–718.

Husain, F. (1998), Seasonality in the Pakistani equity market: the Ramadhan effect, *Pakistan Development Review*, **37**, 77–81.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H. and Lee, T.C. (1988), *Introduction to the Theory and Practice of Econometrics*, John Wiley & Sons, Inc, New York.

Kim, S. and Rui, M. (1999), Price, volume and volatility spillovers among New York, Tokyo and London stock markets, *International Journal of Business*, **4**, 41–61.

Kim, S. and Tsurumi, H. (2000), Korean currency crisis and regime change: a multivariate GARCH model with Bayesian approach, *Asia-Pacific Financial Markets*, **7**, 31–44.

Koutmos, G. (1992), Asymmetric volatility and risk return tradeoff in foreign stock markets, *Journal of Multinational Financial Management*, **2**, 27–43.

Kyle, A.S. (1985), Continuous auctions and insider trading, *Econometrica*, **53**, 1315–1335.

Longin, F.M. (1997), The threshold effect in expected volatility: a model based on asymmetric information, *Review of Financial Studies*, **10**, 837–869.

Loy, J.P. and Weaver, R.D. (1998), Inflation and relative price volatility in Russian food markets, *European Review of Agricultural Economics*, **25**, 373–394.

Lumsdaine, R.L. (1991), Asymptotic properties of the maximum likelihood estimator in GARCH(1, 1) and IGARCH(1, 1) models, *Princeton University Department of Economics manuscript*.

Mecagni, M. and Sourial, M.S. (1999), The Egyptian stock market: efficiency tests and volatility effects, *International Monetary Fund Working Paper*: 99/48.

Mundaca, B.G. (1991), The volatility of the Norwegian currency basket, *Scandinavian Journal of Economics*, **93**, 53–73.

Nas, T.F. and Perry, M.J. (2000), Inflation, inflation uncertainty, and monetary policy in Turkey: 1960–1998, *Contemporary Economic Policy*, **18**, 170–180.

Nelson, D.B. (1990), Stationarity and persistence in the GARCH(1, 1) model, *Econometric Theory*, **6**, 318–334.

Nelson, D.B. (1991), Conditional heteroskedasticity in asset returns: a new approach, *Econometrica*, **59**, 347–370.

Niarchos, N., Tse, Y., Wu, C. and Young, A. (1999), International transmission of information: a study of the relationship between the U.S. and Greek stock markets, *Multinational Finance Journal*, **3**, 19–40.

Ruiz, E. (1994), Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics*, **63**, 289–306.

Shields, K.K. (1997), Threshold modelling of stock return volatility on Eastern European markets, *Economics of Planning*, **30**, 107–125.

Tai, C.S. (1999), Time-varying risk premia in foreign exchange and equity markets: evidence from Asia-Pacific countries, *Journal of Multinational Financial Management*, **9**, 291–316.

Tay, N.S.P. and Zhu, Z. (2000), Correlations in returns and volatilities in Pacific-Rim stock markets, *Open Economies Review*, **11**, 27–47.

Wang, P.J. and Wang, P. (1999), Foreign exchange market volatility in Southeast Asia, *Asia Pacific Financial Markets*, **6**, 235–252.

Wang, P.J. and Wang, P. (2001), Equilibrium adjustment, basis risk and risk transmission in spot and forward foreign exchange markets, *Applied Financial Economics*, **11**, 127–136.

Wang, P.J., Wang, P. and Topham, N. (1999), Relative price variability and inflation uncertainty, *Applied Economics*, **31**, 1531–1539.

6 Shock persistence and impulse response analysis

From the study of unit roots in Chapter 4 we have known the distinctive characteristics of stationary and non-stationary time series. Nevertheless, one of the main concerns in Chapter 4 was whether a time series has a unit root or not, but there was no further examination regarding different properties of non-stationary time series – whether they are a pure random walk or possess serial correlation. Further, what is the serial correlation structure of a time series if it is not a pure random walk? There are generally two categories of non-pure random walk time series. If the time series can be viewed as a combination of a pure random walk process and a stationary process with serial correlation, the long-run effect would be smaller than that of a pure random walk, and the time series contains unit roots due to its non-stationary component. If there is no stationary component in the time series which is not a pure random walk either, then the first difference of the time series is a stationary process with serial correlation, and the long-run effect would be larger that that of a pure random walk. There would be, to a certain degree, mean-reverting tendency in the former category due to its stationary component; and there would be compounding effect in the latter. The interest in this chapter is then centred on the characteristics and behaviour of time series associated with their correlation structure, and relative contribution and importance of the two components: the trend which is a pure random walk, and the cycle (after taking the first difference in the latter category) which is a stationary process involving serial correlation, in the long-run. How persistent is a time series depends on the relative contribution of the two components.

This chapter first discusses measures of persistence in time series in both univariate and multivariate cases. Then the chapter introduces impulse response analysis, which, in a similar way but from a different perspective to persistence analysis, shows graphically the path of response in a time series to a shock to itself or to another time series. Both orthogonal impulse response analysis and non-orthogonal cross-effect impulse response analysis are considered, together with their related and respective variance decomposition.

6.1. Univariate persistence measures

Economic time series are usually a combination of a non-stationary trend component and a stationary cycle component, shocks to the two components are different in that they have remarkably different effects on future trend values. A shock to a stationary time series is transitory and the effect will disappear after a sufficient long time. Using a simple first-order autoregressive process for example:

$$y_{1,t} = c + \rho y_{1,t-1} + \varepsilon_{1,t} \tag{6.1}$$

where $\rho < 1$. Suppose there is a shock at time t with its magnitude being s, and there is no shock afterwards. Then after k periods, the time series evolves to:

$$y_{1,t+k} = \frac{1 - \rho^{k+1}}{1 - \rho} c + \rho^{k+1} y_{1,t-1} + \rho^k s$$

$$\underset{k \to \infty}{=} \frac{c}{1 - \rho} \tag{6.2}$$

i.e. the time series reverts to its mean value and the impact of the shock disappears, the smaller the value of ρ, the quicker. In contrast, a shock to a trend as expressed in a pure random walk moves the time series away from its trend path permanently by an extent which is exactly the size of the shock. For example, if in the following random walk process:

$$y_{2,t} = c + y_{2,t-1} + \varepsilon_{2,t} \tag{6.3}$$

there is a shock at time t with a magnitude of s, and there is no further shock afterwards. Then after k periods the impact is to shift permanently the level of the time series by an extent of s:

$$y_{2,t+k} = (k+1)c + y_{2,t-1} + s \tag{6.4}$$

The impact will not disappear even if $k \to \infty$.

If there is a third time series which is a combination of a stationary time series of the kind of equation (6.1) and a pure random walk such as equation (6.3), then the impact of a shock will not disappear, nor the impact would exactly be s. The permanent impact would usually be a figure smaller than s, depending on the relative contributions of the trend component and cycle component. Furthermore, if $\rho > 1$ in equation (6.1), then the first difference of the time series is stationary and the impact of a shock will disappear after a sufficient long time, while the impact of a shock to the time series itself would be greater than that to a pure random walk. Persistence is therefore introduced as a concept and measure for long-run or permanent impact of shocks on time series, taking above illustrated behaviour and patterns, which are beyond the question of testing for unit roots, into consideration. We first describe persistence with the infinite polynomial of the

Wold moving average representation of time series, as adopted by Campbell and Mankiw (1987a,b); and then introduce more effective methods for its estimation and the ideas behind those methods. Persistence can be illustrated by the infinite polynomial of the Wold moving average representation of a time series, $A(L)$, being evaluated at $L = 1$, i.e.:

$$\Delta Y_t = A(L)\varepsilon_t, \quad \varepsilon_t \sim (0, \sigma_\varepsilon^2) \tag{6.5}$$

where

$$A(L) = 1 + A_1 L + A_2 L^2 + \cdots \tag{6.6}$$

is a polynomial in the lag operator L, ε_t is zero mean and independent (not necessarily iid) residuals. $A(1)(= 1 + A_1 + A_2 + \cdots)$ is $A(L)$ valued at $L = 1$. The impact of a shock in period t on the change or first difference of the time series in period $t + k$ is A_k. The impact of the shock on the level of the time series in period $t + k$ is therefore $1 + A_1 + \cdots + A_k$. The accumulated impact of the shock on the level of the time series is the infinite sum of these moving average coefficients $A(1)$. The value of $A(1)$ can then be used as a measure of persistence. In a pure random walk, $A(1) = 1$; and in any stationary time series, $A(1) = 0$. For series which are neither stationary nor a pure random walk, $A(1)$ can take on any value greater than zero. If $0 < A(1) < 1$, the time series would display mean-reversion tendency. If $A(1) > 1$, an unanticipated increase would be reinforced by other positive changes in the future, and the series would continue to diverge from its pre-shock expected level.

Having introduced the above straightforward representation of persistence, we discuss a second and non-parametric approach to measuring persistence proposed by Cochrane (1988), which is the ratio of the k-period variance to the one period variance, being divided by $k + 1$. The method of the infinite polynomial of the Wold moving average representation involves estimation of parameters $A(L)$ which is sensitive to change. The variance ratio method is non-parametric and the estimate is consequently more stable. The Cochrane (1988) persistence measure is known as V_k in the following formula:

$$V_k = \frac{1}{k+1} \frac{\text{Var}(\Delta_k y_t)}{\text{Var}(\Delta y_t)} = 1 + \frac{2}{\sigma_{\Delta y_t}^2} \sum_\tau^k \left(1 - \frac{\tau}{k+1}\right) \text{Cov}(\Delta y_t, \Delta y_{t-\tau})$$

$$= 1 + \frac{2}{\sigma_{\Delta y_t}^2} \sum_\tau^k \left(1 - \frac{\tau}{k+1}\right) R_\tau = 1 + 2 \sum_\tau^k \left(1 - \frac{\tau}{k+1}\right) \rho_\tau \tag{6.7}$$

where Δ_k is the k-period difference operator and $\Delta_k y_t = y_t - y_{t-k}$; Δ is the usual one period difference operator and the subscript 1 is suppressed for

simplicity, $R_\tau = \text{Cov}(\Delta y_t, \Delta y_{t-\tau})$ is the τth autocovariance in Δy_t, and $\rho_\tau = \text{Cov}(\Delta y_t, \Delta y_{t-\tau})/\sigma^2_{\Delta y_t}$ is the τth autocorrelation in Δy_t. The right-hand side of equation (6.7) is in fact the spectrum of Δy_t at the zero frequency, passing through a k-size window. Interested readers can refer this to Chapter 10 for detail.

In theory, the relationship between V_k and $A(1)$, ignoring any inaccuracies in estimation, is $V_k = A^2(1)(\sigma^2_\varepsilon \sigma^2_{\Delta y_t})$. So let us define persistence consistently as follows:

$$P = V_k = A^2(1)\frac{\sigma^2_\varepsilon}{\sigma^2_{\Delta y_t}} \tag{6.8}$$

But, as one cannot effectively estimate $A(1)$, one cannot effectively estimate P via $A(1)$ either. This is one of the reasons for having a V_k version of persistence. To empirically obtain the persistence measurement, approaches include ARMA, non-parametric, and unobserved components methods. The ARMA approach is to estimate $A(1)$ direct, where parameters are quite sensible to change with regard to estimation. The non-parametric approach is then widely adopted and has been written as two RATS procedures by Goerlich (1992).

In the random walk circumstance, the variance of the k-period difference of a time series is k times the variance of the one period difference of the time series, then the persistence measure $V_k = 1$. For any stationary series, the variance of the k-period difference approaches twice the variance of the one period difference. In this case, V_k approaches zero when k becomes larger. The limit of the ratio of the two variances is therefore the measure of persistence. The choice of k, the number of autocorrelations to be included, is important. Too few autocorrelations may obscure trend-reversion tendency in higher order autocorrelations; and too many autocorrelations may exaggerate the trend-reversion, since as k approaches the sample size T, the estimator approaches zero. Hence, though larger k might be preferred, k must be small relative to the sample size.

6.2. Multivariate persistence measures

The persistence measures of V_k and $A(1)$ can be generalised and applied to multivariate time series. The multivariate V_k and $A(1)$ can then be jointly applied to a group of variables or sectors, e.g. industrial production, construction and services, to evaluate the cross-section effects. Again, we first adopt the infinite polynomial of the Wold moving average representation to demonstrate persistence measures in a way similar to equation (6.5):

$$\Delta \mathbf{y}_t = \mathbf{A}(\mathbf{L})\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim (\mathbf{0}, \boldsymbol{\Sigma}_\varepsilon) \tag{6.9}$$

where we use characters in bold for matrices and vectors.

$$\mathbf{A}(\mathbf{L}) = \mathbf{A}_0 + \mathbf{A}_1\mathbf{L} + \mathbf{A}_2\mathbf{L}^2 + \cdots \tag{6.10}$$

is an $n \times 1$ dimension vector of infinite polynomials, $\Delta \mathbf{y}_t$ is an $n \times 1$ dimension vector of variables, $\boldsymbol{\varepsilon}_t$ is an $n \times 1$ dimension vector of residuals, and $\boldsymbol{\Sigma}_\varepsilon$ is an $n \times n$ covariance matrix of residuals. Similar to the univariate case and extending equation (6.7), we have multivariate persistence measure as follows:

$$\mathbf{P} = \mathbf{A(1)} \boldsymbol{\Sigma}_\varepsilon \boldsymbol{\Sigma}_{\Delta y_t}^{-1} \mathbf{A(1)}' \tag{6.11}$$

which reduces to $A^2(1)(\sigma_\varepsilon^2 / \sigma_{\Delta y_t}^2)$ in a univariate time series.

To obtain multivariate persistence measures, previous studies have attempted to scale the covariance matrix of residuals in different ways. Pesaran *et al.* (1993) use the conditional variance of $\Delta y_{j,t}$ (the jth diagonal element of $\boldsymbol{\Sigma}_\varepsilon$) to normalise the jth column of the covariance matrix of residuals. Van de Gucht *et al.* (1996) use the unconditional variance of $\Delta y_{j,t}$ (the jth diagonal element of $\boldsymbol{\Sigma}_{\Delta y_t}$) to scale the jth column of the covariance matrix of residuals, arguing that it is consistent with the univariate persistence measure proposed by Cochrane (1988). Both Van de Gucht *et al.* (1996) and Pesaran *et al.* (1993) regard the diagonal elements in the normalised covariance matrix as representing total persistence in individual sectors, and off-diagonal elements as the cross effect between two sectors, e.g. an element in the ith row and the jth column is the effect on the ith sector due to a shock in the jth sector. Both aim to generalise the persistence measure and have partly achieved this objective. They have extended the persistence measurement to the multivariate case. However, their normalisations use a single variance for the normalisation of a column and, whether conditional or unconditional, ignore the fact that the process is multivariate. In fact, the normalisation is as simple as in univariate cases. Instead of being scaled down by the unconditional variance, the covariance matrix of residuals should be normalised by the unconditional covariance matrix, i.e. the covariance matrix for $\Delta \mathbf{y}_t$. To have an exact expression of multivariate persistence, the normalisation should be realised with matrix operations; it is not possible to achieve this with the simple dividing arithmetic.

By considering possible effects from, and links with, other sectors, this measurement of multivariate persistence for individual sectors is more precise, compared with its univariate counterpart. With this approach, the effect on sector i due to shocks in sector j is represented by the (i, j) element in $\mathbf{P}$, i.e. $\mathbf{P}(i, j)$, while $\mathbf{P}(i, i)$ measures the sector-specific persistence.

Generalising the non-parametric persistence measure into the multivariate case, we define $\mathbf{V}_k$ as the k-period covariance matrix times the inverse of the one period covariance matrix, divided by $k + 1$:

$$\mathbf{V}_k = \frac{1}{k+1} \boldsymbol{\Sigma}_{\Delta_k y_t} \boldsymbol{\Sigma}_{\Delta y_t}^{-1} \tag{6.12}$$

In a procedure equivalent to equation (6.7), letting $\Delta \mathbf{y}_t$ pass through a k-size window in the Fourier transform and evaluating at the zero

frequency, we have:

$$
\mathbf{V}_k =
\begin{bmatrix}
1 + 2\sum_{\tau}^{k}\left(1 - \frac{\tau}{k+1}\right)R_{11,\tau} & \cdots & \cdots & 1 + 2\sum_{\tau}^{k}\left(1 - \frac{\tau}{k+1}\right)R_{1n,\tau} \\
& \cdots & & \\
1 + 2\sum_{\tau}^{k}\left(1 - \frac{\tau}{k+1}\right)R_{n1,\tau} & \cdots & \cdots & 1 + 2\sum_{\tau}^{k}\left(1 - \frac{\tau}{k+1}\right)R_{nn,\tau}
\end{bmatrix}
$$

$$
\times
\begin{bmatrix}
R_{11,0} & \cdots & \cdots & R_{1n,0} \\
\cdots & & & \\
R_{n1,0} & & & R_{nn,0}
\end{bmatrix}^{-1}
\tag{6.13}
$$

where $R_{ij,\tau} = \mathrm{Cov}(\Delta y_{i,t}, \Delta y_{j,t-\tau})$ is the covariance between $\Delta y_{i,t}$ and $\Delta y_{j,t}$ at lag τ, and $R_{ij,0} = \mathrm{Cov}(\Delta y_{i,t}, \Delta y_{j,t})$ is the contemporaneous covariance. The elements in the first matrix on the right-hand side are bivariate, but the elements in $\mathbf{V}_k$ are truly multivariate due to the second matrix on the right-hand side. So this measure of persistence takes account of the influence from all the sources in the system when considering $\mathbf{V}_k(i, j)$ in the appearance of interactions between the ith and jth time series.

Multivariate persistence analysis is more sensible in that, instead of analysing the individual variables separately as in univariate cases, it allows shocks to transmit from one variable to all the others. Therefore, multivariate persistence analysis is able to examine the sources of shocks and the effects of the shock in one sector on other sectors. Moreover, it is able to detect the effect of certain kinds of shocks, e.g. a monetary shock, from that of other shocks, e.g. shocks from the real sectors. The multivariate measurement of persistence is not built on structural relations. As such, the inclusion of a specific kind of shock in persistence analysis will not lead to the violation of constraints, as it may in a system of structural equations. In addition, the effects of this specific shock can be evaluated in a VAR framework, which is relatively less complicated. A specific kind of shock can be added to the model, as in the following:

$$
\Delta \mathbf{y}_t = \mathbf{s(L)}\mathbf{v}_t + \mathbf{A(L)}\boldsymbol{\varepsilon}_t
\tag{6.14}
$$

where $\mathbf{v}_t$ represents specific shocks whose effects are to be analysed, which can be the demand shock, supply shock or monetary shock, depending on the way it is extracted from another fitted equation(s); and $\mathbf{s(L)}$ is an $n \times 1$ dimension vector of polynomials. By evaluating equation (6.14) with and without $\mathbf{v}_t$, one can establish whether an individual sector is subject to shock $\mathbf{v}_t$. Furthermore, in the existence of the effect of $\mathbf{v}_t$, the proportion of the persistence due to $\mathbf{v}_t$ and that of other shocks can be identified. In theory, more than one set of specific shocks can be included; in which case, $\mathbf{v}_t$ becomes an m dimension vector with m being the number of sets of shocks, and $\mathbf{s(L)}$ is an $n \times m$ matrix. However, the estimation would be empirically unfeasible, since greater inaccuracy would be introduced.

In addition, this approach would be less appealing if it is to lose its advantages of no subjective assumptions and restrictions. Nevertheless, if there are only two types of shocks, e.g. demand versus supply, or monetary versus real, then $\boldsymbol{\nu}_t$ can only be one of the two types of shocks, otherwise equation (6.14) would be over-identified.

Persistence can be decomposed into separate components due to the specific shock and that due to other shocks:

$$\mathbf{P}_s = \mathbf{A}(1)\mathbf{s}(1)\boldsymbol{\Sigma}_\nu^2 \mathbf{s}(1)'\boldsymbol{\Sigma}_{\Delta y_t}^{-1}\mathbf{A}(1)' \tag{6.15}$$

$$\mathbf{P}_o = \mathbf{A}(1)\boldsymbol{\Sigma}_\varepsilon \boldsymbol{\Sigma}_{\Delta y_t}^{-1}\mathbf{A}(1)' \tag{6.16}$$

and total persistence is:

$$\mathbf{P}_T = \mathbf{P}_s + \mathbf{P}_o \tag{6.17}$$

If the specific shock is chosen as demand or monetary disturbance, then the underlying assumption is that the demand or monetary shock may also have a long-run effect, as the persistence measure is about the effect on the levels of variables. This assumption can be empirically ruled out or ruled in which, in fact, becomes a hypothesis. Although Blanchard and Quah (1989) arguably excluded the demand shock from having a long-run effect, their empirical work suggests that the effect of a demand shock would decline to vanish in about 25 quarters or five to six years. In such a long period, the probability of a structural change or break would be rather high. If a structural change does happen, it would override any supply shocks and the effects of demand and supply shocks would be mixed.

6.3. Impulse response analysis and variance decomposition

Impulse response analysis is another way of inspecting and evaluating the impact of shocks cross-section. While persistence measures focus on the long-run properties of shocks, impulse response traces the evolutionary path of the impact over time. Impulse response, together with variance decomposition, forms innovation accounting for sources of information and information transmission in a multivariate dynamic system.

Considering the following VAR process:

$$\mathbf{y}_t = \mathbf{A}_0 + \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2} + \cdots + \mathbf{A}_k \mathbf{y}_{t-k} + \boldsymbol{\mu}_t \tag{6.18}$$

where $\mathbf{y}_t$ is an $n \times 1$ vector of variables, $\mathbf{A}_0$ is an $n \times 1$ vector of intercept, $\mathbf{A}_\tau (\tau = 1, \ldots, k)$ are $n \times n$ matrices of coefficients, $\boldsymbol{\mu}_t$ is an n dimension vector of white noise processes with $E(\boldsymbol{\mu}_t) = \mathbf{0}$, $\boldsymbol{\Sigma}_\mu = E(\boldsymbol{\mu}_t \boldsymbol{\mu}_t')$ being non-singular for all t, and $E(\boldsymbol{\mu}_t \boldsymbol{\mu}_s')$ for $t \neq s$. Without losing generality, exogenous variables other than

lagged $\mathbf{y}_t$ are omitted for simplicity. A stationary VAR process of equation (6.18) can be shown to have a moving average (MA) representation of the following:

$$\mathbf{y}_t = \mathbf{C} + \boldsymbol{\mu}_t + \boldsymbol{\Phi}_1 \boldsymbol{\mu}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{\mu}_{t-2} + \cdots$$

$$= \mathbf{C} + \sum_{\tau=0}^{\infty} \boldsymbol{\Phi}_\tau \boldsymbol{\mu}_{t-\tau} \tag{6.19}$$

where $\mathbf{C} = E(\mathbf{y}_t) = (\mathbf{I} - \mathbf{A}_1 - \cdots - \mathbf{A}_k)^{-1} \mathbf{A}_0$, and $\boldsymbol{\Phi}_\tau$ can be computed from $\mathbf{A}_\tau$ recursively $\boldsymbol{\Phi}_\tau = \mathbf{A}_1 \boldsymbol{\Phi}_{\tau-1} + \mathbf{A}_2 \boldsymbol{\Phi}_{\tau-2} + \cdots + \mathbf{A}_k \boldsymbol{\Phi}_{\tau-k}$, $\tau = 1, 2, \ldots$, with $\boldsymbol{\Phi}_0 = \mathbf{I}$ and $\boldsymbol{\Phi}_\tau = \mathbf{0}$ for $\tau < 0$.

The MA coefficients in equation (6.19) can be used to examine the interaction between variables. For example, $a_{ij,k}$, the ijth element of $\boldsymbol{\Phi}_k$, is interpreted as the reaction, or impulse response, of the ith variable to a shock τ periods ago in the jth variable, provided that the effect is isolated from the influence of other shocks in the system. So a seemingly crucial problem in the study of impulse response is to isolate the effect of a shock on a variable of interest from the influence of all other shocks, which is achieved mainly through orthogonalisation.

Orthogonalisation *per se* is straightforward and simple. The covariance matrix $\boldsymbol{\Sigma}_\mu = E(\boldsymbol{\mu}_t \boldsymbol{\mu}_t')$, in general, has non-zero off-diagonal elements. Orthogonalisation is a transformation, which results in a set of new residuals or innovations $\mathbf{v}_t$ satisfying $E(\mathbf{v}_t \mathbf{v}_t') = \mathbf{I}$. The procedure is to choose any non-singular matrix $\mathbf{G}$ of transformation for $\mathbf{v}_t = \mathbf{G}^{-1} \boldsymbol{\mu}_t$ so that $\mathbf{G}^{-1} \boldsymbol{\Sigma}_\mu \mathbf{G}^{-1} = \mathbf{I}$. In the process of transformation or orthogonalisation, $\boldsymbol{\Phi}_\tau$ is replaced by $\boldsymbol{\Phi}_\tau \mathbf{G}$ and $\boldsymbol{\mu}_t$ is replaced by $\mathbf{v}_t = \mathbf{G}^{-1} \boldsymbol{\mu}_t$, and equation (6.19) becomes:

$$\mathbf{y}_t = \mathbf{C} + \sum_{\tau=0}^{\infty} \boldsymbol{\Phi}_\tau \boldsymbol{\mu}_{t-\tau} = \mathbf{C} + \sum_{\tau=0}^{\infty} \boldsymbol{\Phi}_\tau \mathbf{G} \mathbf{v}_{t-\tau}, \quad E(\mathbf{v}_t \mathbf{v}_t') = \mathbf{I} \tag{6.20}$$

Suppose that there is a unit shock to, for example, the jth variable at time 0 and there is no further shock afterwards, and there are no shocks to any other variables. Then after k periods $\mathbf{y}_t$ will evolve to the level:

$$\mathbf{y}_{t+k} = \mathbf{C} + \left(\sum_{\tau=0}^{k} \boldsymbol{\Phi}_\tau \mathbf{G} \right) \mathbf{e}(\mathbf{j}) \tag{6.21}$$

where $\mathbf{e}(\mathbf{j})$ is a selecting vector with its jth element being one and all other elements being zero. The accumulated impact is the summation of the coefficient matrices from time 0 to k. This is made possible because the covariance matrix of the transformed residuals is a unit matrix $\mathbf{I}$ with off-diagonal elements being zero. Impulse response is usually exhibited graphically based on equation (6.21). A shock to each of the n variables in the system results in n impulse response functions and graphs, so there are a total of $n \times n$ graphs showing these impulse response functions.

To achieve orthogonalisation, the Choleski factorisation, which decomposes the covariance matrix of residuals Σ_μ into $\mathbf{GG}'$ so that $\mathbf{G}$ is lower triangular with positive diagonal elements, is commonly used. However, this approach is not invariant to the ordering of the variables in the system. In choosing the ordering of the variables, one may consider their statistical characteristics. By construction of $\mathbf{G}$, the first variable in the ordering explains all of its one-step forecast variance, so consign a variable which is least influenced by other variables, such as an exogenous variable, to the first in the ordering. Then choose the variable with least influence on other variables as the last variable in the ordering. The other approach to orthogonalisation is based on the economic attributes of data, such as the Blanchard and Quah structural decomposition. It is assumed that there are two types of shocks, the supply shock and the demand shock. While the supply shock has permanent effect, the demand shock has only temporary or transitory effect. Restrictions are imposed accordingly to realise orthogonalisation in residuals.

Since the residuals have been orthogonalised, variance decomposition is straightforward. The k-period ahead forecast errors in equation (6.19) or (6.20) are:

$$\sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{G} \mathbf{v}_{t-\tau+k-1} \tag{6.22}$$

The covariance matrix of the k-period ahead forecast errors are:

$$\sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{G} \mathbf{G}' \mathbf{\Phi}_\tau' = \sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \Sigma_\mu \mathbf{\Phi}_\tau' \tag{6.23}$$

The right-hand side of equation (6.23) just reminds the reader that the outcome of variance decomposition will be the same irrespective of $\mathbf{G}$. The choice or derivation of matrix $\mathbf{G}$ only matters when the impulse response function is concerned to isolate the effect from the influence from other sources.

The variance of forecast errors attributed to a shock to the jth variable can be picked out by a selecting vector $\mathbf{e(j)}$, the idea of variance decomposition, with the jth element being one and all other elements being zero:

$$\text{Var}(\mathbf{j}, \mathbf{k}) = \left(\sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{G} \mathbf{e(j)} \mathbf{e(j)}' \mathbf{G}' \mathbf{\Phi}_\tau' \right) \tag{6.24}$$

Further, the effect on the ith variable due to a shock to the jth variable, or the contribution to the ith variable's forecast error by a shock to the jth variable, can be picked out by a second selecting vector $\mathbf{e(i)}$ with the ith element being one and all other elements being zero.

$$\text{Var}(ij, k) = \mathbf{e(i)}' \left(\sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{G} \mathbf{e(j)} \mathbf{e(j)}' \mathbf{G}' \mathbf{\Phi}_\tau' \right) \mathbf{e(i)} \tag{6.25}$$

In relative terms, the contribution is expressed as a percentage of the total variance:

$$\frac{\text{Var}(ij, k)}{\sum_{j=1}^{n} \text{Var}(ij, k)} \tag{6.26}$$

which sums up to 100 per cent.

6.4. Non-orthogonal cross-effect impulse response analysis

There are other ways to evaluate the effect of a shock. One of the main advantages of applying orthogonalised residuals is that the impact at time k due to a unit shock to the jth variable at time 0 is simply the summation of matrices $\mathbf{\Phi}_\tau \mathbf{G}$, over $0 \le \tau \le k$, being timed by the selecting vector $\mathbf{e(j)}$. That is, there is no need to consider the effect due to shocks to other than the jth variable because such effect does not exist. Then it would be a reasonable idea that we do not perform orthogonalisation but consider the effect arising from the non-orthogonalisation of residuals, or the cross-effect, in impulse response analysis. With non-orthogonal residuals, when there is a shock to the jth variable of the size of its standard deviation, there are shocks to other variables in the meantime through their correlations. Let $\boldsymbol{\delta}_j$ stand for such shocks:

$$\boldsymbol{\delta}_j = \begin{bmatrix} \rho_{1j} \\ \cdots \\ 1 \\ \rho_{nj} \end{bmatrix} \sqrt{\sigma_{jj}} = \begin{bmatrix} \sigma_{1j} \\ \cdots \\ \sigma_{jj} \\ \sigma_{nj} \end{bmatrix} \frac{1}{\sqrt{\sigma_{jj}}} = \mathbf{\Sigma}_\mu \mathbf{e(j)} \frac{1}{\sqrt{\sigma_{jj}}} \tag{6.27}$$

With such a shock to the jth variable at time 0 and suppose there is no further shock afterwards, $\mathbf{y}_t$ will evolve to the level after k periods:

$$\mathbf{y}_{t+k} = \mathbf{C} + \left(\sum_{\tau=0}^{k} \mathbf{\Phi}_\tau \right) \mathbf{e(j)} \frac{1}{\sqrt{\sigma_{jj}}} \tag{6.28}$$

So, it appears that orthogonalisation can be avoided. But bearing in mind that, in non-orthogonal impulse response analysis, we cannot simply give a shock of one standard deviation to the equation of interest, the jth equation, only; we should, in the meantime, give a 'shock' to each of other equations of the size of the square root of its covariance with the jth shock. It indeed means that we have to consider both the direct effect of the jth shock and the indirect effect of the jth shock through other series in the system. Moreover, the outcome would be in general different to that from orthogonal impulse response analysis.

Let us work out variance decomposition in a slightly different way. We consider single elements first. The k-period ahead forecast errors in equation (6.28) are:

$$\sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{\Sigma}_\mu \mathbf{e(j)} \frac{1}{\sqrt{\sigma_{jj}}} \tag{6.29}$$

The covariance matrix of the k-period ahead forecast errors contributed to the jth shock are:

$$\text{Var}(\mathbf{j}, \mathbf{k}) = \frac{1}{\sigma_{jj}} \sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{\Sigma}_\mu \mathbf{e}(\mathbf{j})\mathbf{e}(\mathbf{j})' \mathbf{\Sigma}_\mu \mathbf{\Phi}_\tau' \tag{6.30}$$

The total covariance matrix is the summation of equation (6.30) over j:

$$\sum_{j=1}^{n} \frac{1}{\sigma_{jj}} \sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{\Sigma}_\mu \mathbf{e}(\mathbf{j})\mathbf{e}(\mathbf{j})' \mathbf{\Sigma}_\mu \mathbf{\Phi}_\tau' \tag{6.31}$$

which is different from equation (6.23). The variance of the ith variable contributed to the jth shock and the total variance of the ith variable are:

$$\text{Var}(ij, k) = \frac{1}{\sigma_{jj}} \mathbf{e}(\mathbf{i})' \left(\sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{\Sigma}_\mu \mathbf{e}(\mathbf{j})\mathbf{e}(\mathbf{j})' \mathbf{\Sigma}_\mu \mathbf{\Phi}_\tau' \right) \mathbf{e}(\mathbf{i}) \tag{6.32}$$

and:

$$\text{Var}(i, k) = \sum_{j=1}^{n} \frac{1}{\sigma_{jj}} \mathbf{e}(\mathbf{i})' \left(\sum_{\tau=0}^{k-1} \mathbf{\Phi}_\tau \mathbf{\Sigma}_\mu \mathbf{e}(\mathbf{j})\mathbf{e}(\mathbf{j})' \mathbf{\Sigma}_\mu \mathbf{\Phi}_\tau' \right) \mathbf{e}(\mathbf{i}) \tag{6.33}$$

respectively. The contribution by the jth shock expressed as a percentage of the total variance is:

$$\frac{\text{Var}(ij, k)}{\text{Var}(j, k)} = \frac{\text{Var}(ij, k)}{\sum_{j=1}^{n} \text{Var}(ij, k)} \tag{6.34}$$

which sums up to 100 per cent.[1]

6.5. Examples and cases

Example 6.1

This case presents the profile of the UK property market's responses to shocks of various sources by Wang (2000). We only use and discuss the multivariate part of the study. The variables considered in the study in relation to persistence in the property market are the Jones Lang Wootten property total return index (JLW) (with necessary adjustment), the Nationwide Building Society House Price Index (NTW), the Financial Times Actuary All Share Index (FTA), Construction output on new work (CO), Total production (PDN), Services (SVC), the

Continued

Unemployment rate (UER), and Money supply (M0). All the economic data are of quarterly frequency and from the Office for National Statistics (ONS) of the UK. They are all seasonally adjusted for consistency, as not all data are available in the form of non-seasonally adjusted. Table 6.1 presents the multivariate persistence estimates with the six sectors represented by JLW, FTAP, NWT, CO, PDN, and SVC. The diagonal elements in the table are sector-specific persistence measures (the diagonal elements in the **P** or $\mathbf{V}_k$ matrix). FTA, the stock market index, is most close to a random walk with its $A(1)k$ being very close to unity. Total production and services do not have as large persistence measures as property, housing and construction. No direct comparison with other studies is possible, because there have been virtually no studies of persistence of shocks in the UK economy and sectors. Several US studies have reported that the services sector has a large persistence measure estimate while the production sector has a relatively low value for persistence measurement. It is also documented that utilities exhibit considerable persistence while manufacturing has a rather small value for persistence measurement. In Table 6.1 the production sector's V_k of 1.3348 would be an aggregate estimate combining a higher value of persistence for utilities and lower value of persistence for manufacturing. As the intention of this multivariate persistence analysis is to investigate the cross-sectional effects between property and the broadly classified sectors, no further disaggregation is necessary and appropriate here.

The off-diagonal elements in Table 6.1 provide information that is not found in univariate persistence analysis. It has been revealed that shocks from the housing market have the largest effect on the persistence in property, with the cross-sectional effect on JLW from NTW being 2.7542. It is followed by the services sector which is also quite substantial, the

Table 6.1 Multivariate persistence

| Effect on | Sources of shocks | | | | | |
	JLW	FTA	NTW	CO	PDN	SVC
JLW	2.6243	0.8744	2.7542	1.0119	1.0782	1.5273
FTA	0.7962	0.8284	1.2710	0.6919	0.1238	1.0530
NTW	3.0412	1.5032	4.4957	1.3024	1.3020	2.3466
CO	−0.3611	−0.0746	−0.9907	2.8815	0.7135	0.8976
PDN	0.4385	−0.1939	0.0468	1.5431	1.3348	0.5788
SVC	0.8457	0.6488	0.9700	1.9384	0.7768	1.5798

Note: Same as in the univariate cases, the standard error of these statistics is $4(k+1)/3NVC_k(i,j)$ (with the Bartlett window), where N is the number of observations. k is in fact the window size in the frequency domain. With our specification it can be inferred the window size is about 1/4 of the total observations, so the standard error of these statistics is acceptable. See Priestley (1996). Detail from the authors upon request.

production sector, and construction. Shocks in the stock market have effects on the persistence in property, but they are the smallest among all selected variables, with the cross-sectional effect on JLW being 0.8744.

Regarding the effects of the property market on other sectors, again, the largest impacts seem to be felt in the housing market, with the cross-sectional effect on NTW from JLW being 3.0412. So the commercial and non-commercial property markets have very close links in this perspective. The effects of shocks on the services sector (0.8457) are larger than those on the production sector (0.4385), as expected. A negative figure for the effects on construction suggests, in statistical terms, that the one period covariance and the n ($n \rightarrow \infty$) period covariance have different signs. This is only possible in covariance but not in variance. The empirical meaning of a negative cross-sectional persistence measure would be: a positive shock in the property market which also results in an increase in construction (i.e. a positive one period covariance is assumed) would eventually lead to a decrease in construction output, or contraction in the construction industry, in the long-run. This revelation of the interaction between the property market and the construction sector has profound economic implications.

The reported multivariate persistence measurement estimates are derived with an unrestricted VAR model of order 2 (the inverse of the matrix is effectively corresponding to an infinite moving average process). The restricted model, which drops the regressors whose t-statistic of coefficient is less than one, is also tested. The two sets of results are similar, so the unrestricted model is adopted for reasons that it is easy to implement in the future and in slightly different situations.[2] This is consistent with Cochrane's (1988) recommendation of including all autocorrelation terms even if they are insignificant. Both models are estimated with SUR (Seemingly Unrelated Regression), though there are no efficiency gains from using an OLS procedure to applying SUR in the unrestricted model.

The paper further decomposes shocks into monetary and non-monetary components. The above tests have analysed the 'sources' of shocks, and the sources are sectors. In the following, the sources are divided into monetary and non-monetary ones. The reasons for adopting this line of research are as follows. Traditionally, the effect of a monetary shock is viewed as only being temporary or transitory, while a real shock has both permanent and transitory effects. In the long-run, the effect of the monetary shock disappears and the only effect left is due to the real shock. Similarly, a demand shock is viewed as temporary and a supply shock as permanent. In separating or decomposing a monetary shock from non-monetary shocks, one is able to evaluate the long-run and short-term effects more effectively. However, the traditional view, which suggests that the monetary shock is not held responsible for

Continued

any permanent or long-run effects, may be over-assertive and should be empirically tested. In this study, if the effect of the monetary shock is not long lasting, then the monetary shock would have no contribution to the persistence measure. Obviously, different test results would have different implications for policy-making and practice, especially with regard to a long-run perspective.

Monetary shocks can be derived from estimating a money supply growth model and obtaining its residuals. The money supply growth model is specified as follows:

$$\Delta M_t = \alpha + \beta \Delta M_{t-1} + \gamma \Delta SVC_{t-1} + \delta UER_{t-1} + v_t \tag{6.35}$$

where M_t is money supply, SVC_t is services output and UER_t is the unemployment rate. M0, the narrowly defined money, is chosen as the money supply variable in this model. The reasons for using M0 instead of M4, the broad money supply, are empirical. There is a big break in the M4 series in the fourth quarter of 1981 caused by the switch between the old banking sector and the new monetary sector. In July 1989, Abbey National's conversion to a public limited company caused minor breaks to the M0 series and major breaks in the M4 series. Although the first breaks in the fourth quarter of 1981 were removed from the changes in M4, the removal of the breaks in the changes in M4 resulted in as much distortion as the retaining of the breaks in M4 levels. Besides these breaks, the M0 and M4 series had a similar pattern. Beyond the concern in breaks, M0 is more liquid and more public sensitive in representing demand factors, separated from supply factors or real factors. Table 6.2 reports the summary statistics for the money growth model.

The multivariate shock persistence model has been re-estimated with monetary shocks, the residuals from the money supply growth model, being included. All the estimates are reported in Table 6.3, and a summary with sector-specific estimates and the percentage of monetary and non-monetary effects is provided in Table 6.4. The first line for each variable in Table 6.3

Table 6.2 Summary statistics for the money growth model

	α	β	γ	δ	Q
M0	0.0155***	0.4551***	0.3616***	-0.0011**	19.9744
	(3.2723)	(4.1013)	(2.6350)	(2.4180)	(0.2753)

Q-Ljung–Box statistic for serial correlation, the order is selected as 1/4 of the observations used.
p-Value in parentheses.
**significant at 5 per cent level; ***significant at 1 per cent level.

Table 6.3 Multivariate persistence: monetary shocks decomposed

Effect on	Sources of shocks					
	JLW	FTA	NTW	CO	PDN	SVC
JLW	2.2304	0.4559	2.0688	0.6253	0.6802	0.8212
	2.0389	0.3347	1.5680	0.7949	0.7011	0.7515
	0.1915	0.1212	0.5008	−0.1695	−0.0208	0.0697
FTA	0.3265	0.5301	0.5480	0.4431	−0.1140	0.6191
	0.3253	0.4216	0.3856	0.4922	−0.0267	0.5090
	0.0012	0.1084	0.1624	−0.0492	−0.0874	0.1101
NTW	2.2713	0.7616	3.3395	0.5288	0.7238	1.1391
	1.9208	0.5440	2.4043	0.8496	0.7536	1.0188
	0.3505	0.2176	0.9352	−0.3209	−0.0298	0.1204
CO	−0.4758	−0.0522	−1.2515	3.0167	0.7594	0.9497
	−0.3522	−0.0362	−1.0157	2.9111	0.7699	0.9403
	−0.1236	−0.0160	0.0849	0.1057	−0.0105	0.0094
PDN	0.1860	−0.2489	−0.2548	1.4501	1.2941	0.4737
	0.1776	−0.2774	−0.3397	1.4776	1.2826	0.4481
	0.0083	0.0285	0.0849	−0.0275	0.0115	0.0256
SVC	0.1568	0.3699	−0.0256	1.6847	0.4481	1.2115
	0.2742	0.2303	−0.0559	1.7202	0.6274	1.0124
	−0.1174	0.1396	0.0303	−0.0354	−0.1794	0.1991

Table 6.4 Multivariate persistence: summary of monetary and non-monetary shocks

	Monetary shocks		Non-monetary shocks		
	V_K	%	V_K	%	Total
JLW	0.1915	8.59	2.0389	91.41	2.2304
FTA	0.1084	20.45	0.4216	79.55	0.5301
NTW	0.9352	28.00	2.4043	72.00	3.3395
CO	0.1057	3.50	2.9111	96.50	3.0167
PDN	0.0115	0.89	1.2826	99.11	1.2941
SVC	0.1991	16.43	1.0124	83.57	1.2115

is the total persistence, the second line the effects of non-monetary shocks, as represented by the second term on the right-hand side of equation (6.17), and the third line the effects of monetary shocks represented by the first term on the right-hand side of equation (6.17). As above, the diagonal elements are sector-specific persistence measurement, and off-diagonal elements the cross persistence measurement. Overall, the persistence estimates are smaller than those in Table 6.1, except for the construction sector. This is

Continued

because of the inclusion of the monetary shocks, which are expected to have smaller effects in the long run, in the model. In previous estimation without an explicit monetary shock variable (or a monetary variable), the persistence effects due to monetary shocks are mixed with other shocks. Further scrutiny has found that the decrease in the persistence measure happens in those sectors which are subject to monetary shocks to a substantial degree, e.g. housing, where monetary shocks account for 28 per cent in total persistence, services 16 per cent, and stock market 20 per cent. Monetary shocks only account for 4 per cent of total persistence in construction, and an even smaller figure of less than 1 per cent in the production sector, so their total persistence estimates are largely unaffected. In summary, a broadly defined production sector, including construction, or the real economy, or the supply side of economy, is not subject to monetary shocks in the long-run; whereas the services sector, broadly defined to include housing and the stock market, or the demand side of economy, or the consumption, is very much influenced by monetary shocks. Commercial property, due to its fundamental links to the real economy and financial markets, reasonably stands in between, with the effects of monetary shocks being responsible for 9 per cent of total persistence measurement, and a large part of persistence is from non-monetary shocks caused in the real sector of the economy.

Example 6.2

In a recent paper, Dekker *et al*. (2001) apply both orthogonal and non-orthogonal cross-effect, or generalised, impulse response analysis to stock market linkages in Asia-Pacific. They use daily closing data of returns for a rather short period from the 1st January 1987 to the 29th May 1998, on ten market indices in the region, namely, Australia's SE All Ordinary, Hong Kong Hang Seng, Japan's Nikkei 225 Average, Malaysia's Kuala Lumpur Composite, New Zealand SE Capital 40, the Philippines SE Composite, Singapore Strait Times Industrial, Taiwan SE Weighted, Thailand's Bangkok Book Club, and the US Standard & Poor 500 Composite. Their models were tested using the indices as expressed in the US dollar as well as in local currencies. It is claimed that both data sets produce consistent results so only the results from using local currencies are reported in the paper. Amongst the ten economies, Malaysia, the Philippines, Taiwan and Thailand are classified as emerging markets and the rest as developed markets.

   The models and the treatment of variables are exactly those in Pesaran and Shin (1998). Consequently, variance decomposition with the generalised impulse response procedure inevitably runs into the problem that the total variance does not sum to 100 per cent. The paper deals with the problem by standardising the total variance, or scaling the total variance

to 100 per cent. Although cointegration relationships are found in the data, the authors choose to apply an unrestricted VAR in the first difference without incorporating the error correction term, having reviewed the relevant literature in which an unrestricted VAR is preferred to a vector error correction model (VECM) in short horizons. The paper performs impulse response analysis over 15 days and presents 5-, 10- and 15-day ahead forecast variance decomposition. In orthogonal response analysis, the variables are ordered according to the closing time, with the most exogenous market, which in this case is the US, being the first. Table 6.5 presents the results from orthogonal variance decomposition, while Table 6.6 is for those from generalised variance decomposition. As there is no substantial variation, only the results for day 15 are provided.

The paper makes common sense comparison between the orthogonal and the generalised variance decomposition results. For example, with closing time ordering in orthogonal variance decomposition, New Zealand is ordered before Australia. The ordering appears to have a distorting effect on the variance decomposition results: shocks in the New Zealand market explain a much larger proportion of variance of 10.70 per cent in the Australian market, compared with a rather small figure of 1.99 per cent contributed by the Australian market to the New Zealand market, on day 15. This seems to be difficult to justify, considering the relative size of the two markets. In contrast, generalised variance decomposition provides apparently reasonable results that the contribution of the New Zealand market to the Australian market is 8.31 per cent, while shocks in the Australian market account for a large amount of 11.43 per cent of the total variance in the New Zealand market, on day 15. Following this common sense discussion, the paper employs Table 6.6 for further analysis. There are three main conclusions. First, the US market is the most influential in Asia-Pacific. No other markets contribute more than 2 per cent of the US total forecast variance, while the contribution of the US market to other markets is significant with many of them being over 10 per cent. Second, the level of exogeneity of a market is proportional to the amount of the forecast variance explained by the market itself. The US, with over 90 per cent total forecast variance being accounted for by itself, is the most exogenous. While Singapore is the most endogenous because over 50 per cent total forecast variance is attributed to shocks in the other markets. Third, markets with strong economic ties and close geographic links, such as the pairs of Australia and New Zealand and Malaysia and Singapore, have significant interaction with each other. Impulse response graphs conform the above results. Impulse analysis also indicates that the impact of shocks disappears quickly, usually in no more than one day.

Continued

Table 6.5 Orthogonal decomposition of forecast error variances for daily market returns for 10 Asia Pacific markets: 15 day horizon

Effect on	Innovations in										
	Australia	Hong Kong	Japan	Malaysia	New Zealand	Philippines	Singapore	Taiwan	Thailand	US	All foreign
Australia	61.86	1.2434	0.4190	0.7736	10.70	0.9937	0.3472	0.5010	0.1912	22.95	38.14
Hong Kong	3.7954	76.76	2.2084	0.5356	2.5652	2.4665	0.4848	0.6669	0.6602	9.8596	23.24
Japan	2.4033	0.3245	86.38	0.1338	1.9109	0.3719	0.4858	1.0619	0.2365	6.6953	13.62
Malaysia	1.4956	8.0847	1.6725	54.56	1.6234	1.8302	19.20	0.9947	0.7882	9.7527	45.44
New Zealand	2.4478	1.0501	0.5868	0.4381	77.92	0.4402	0.6593	0.3093	0.2064	15.94	22.08
Philippines	0.2580	0.7532	0.6074	1.2595	0.8184	89.93	0.4930	0.9431	0.7328	4.2023	10.07
Singapore	3.9231	9.8790	1.5353	1.0009	2.1997	3.6799	60.75	1.1406	0.6726	15.22	39.25
Taiwan	0.6490	0.3131	0.1535	0.1467	0.8570	0.2967	0.4565	94.41	0.3786	2.3381	5.59
Thailand	0.7661	2.7976	0.7318	1.6507	0.7751	2.1737	3.9772	1.1219	80.13	5.8802	19.87
US	0.8925	0.6491	0.2373	0.2788	0.3792	0.5864	0.4700	0.3024	0.6313	95.57	4.43

Table 6.6 Generalised decomposition of forecast error variances for daily market returns for 10 Asia Pacific markets: 15 day horizon

| | Innovations in | | | | | | | | | | |
Effect on	Australia	Hong Kong	Japan	Malaysia	New Zealand	Philippines	Singapore	Taiwan	Thailand	US	All foreign
Australia	56.45	5.6008	2.7171	2.1190	8.3087	1.1327	4.5123	0.7183	0.6586	17.78	43.55
Hong Kong	4.3553	60.87	2.7598	7.6287	1.7912	2.1004	11.01	0.6114	2.0003	6.8775	39.13
Japan	3.2464	3.5870	77.98	2.3550	1.6324	0.3189	3.3952	1.0899	0.7011	5.6973	22.03
Malaysia	1.9241	8.2482	1.8899	55.01	1.0477	1.4208	20.18	0.7355	3.2695	6.2784	44.99
New Zealand	11.43	3.5104	2.3341	1.1643	63.81	0.5837	2.6339	0.6866	0.5234	13.05	36.19
Philippines	1.1852	3.8328	0.9343	3.9294	0.7017	77.66	5.3348	0.8805	1.9679	3.5829	22.35
Singapore	3.7789	10.43	2.1103	17.71	1.3428	2.6520	49.01	0.8338	2.9020	9.2371	50.98
Taiwan	1.5537	1.3017	1.2884	1.2419	0.7931	0.5862	2.1089	87.80	1.1669	2.1549	12.20
Thailand	1.1120	4.3221	1.0063	6.1836	0.6352	2.1180	7.5476	1.0419	71.28	4.7574	28.72
US	0.8076	1.9336	1.3817	0.8446	0.3651	0.5874	1.2743	0.3742	0.9412	91.49	8.51

6.6. Empirical literature

Persistence and impulse response are much an empirical matter. Persistence looks into the long-run behaviour of time series in response to shocks and reflects the relative contribution and importance of the trend and the cycle. Inspecting the persistence profile of a time series, the effect of shocks in the long-run can be evaluated, which is of help to both macro economic policy formation and micro investment decision making. The other aspects in the study of the effect of shocks are the response profile over the whole time horizon of interest, including the magnitudes of the response, termed as impulse response analysis, and the examination of the sources of the disturbance, termed as variance decomposition. Many multivariate models, such as the VAR, are complemented with impulse response analysis and variance decomposition, after the model has been set up and tested. All of these reflect the importance of this chapter in empirical studies.

Following the initiatives of Campbell and Mankiw (1987a, b) and Cochrane (1988), whose concerns are the behaviour of US aggregate GNP/GDP data, persistence in macroeconomic time series have been further investigated in the sectors, in other economies and in other economic and financial variables. Pesaran *et al.* (1993) extend measures of persistence into multivariate cases and examine the persistence profile in ten US GNP sectors, though they do not consider the cross-effect of persistence between sectors. Most of the sectors are found to be very persistent in response to shocks, with the persistence measure being greater than one, suggesting there is compounding effect. In comparison, utilities exhibit largest compounding persistence followed by services, while persistence in manufacturing is relatively lower. Mayadunne *et al.* (1995) have carried out similar research using the Australian data and made comparison with the US results. Concerned with the random walk hypothesis in foreign exchange rates, Van de Gucht *et al.* (1996) examine persistence in seven daily foreign spot exchange rates of the Canadian dollar, the French franc, the Swiss franc, the German mark, the Italian lire, the Japanese yen and the British pound vis-à-vis the US dollar over the period of 3 September 1974 to 27 May 1992. They find departure from the random walk benchmark but the departure is not substantial when the standard errors in the persistence measure are taken into consideration. Moreover, there is increasing mean-reverting component in more recent periods. The cross-effect of shocks is also checked and that between European currencies is found to be similar; further, the cross-effect between European currencies is larger than that between European currencies and that of the Japanese yen and the Canadian dollar. Cashin *et al.* (2000) study the persistence of shocks to world commodity prices, using monthly IMF data on primary commodities between 1957 and 1998. They find that shocks to commodity prices typically have significantly persistent effect and the persistence profile varies, based on which the effect of national and international schemes of earnings stabilisation may be formed and evaluated. Their analysis is not in favour of a stabilisation scheme, as they argue that the cost of the stabilisation scheme will likely exceed any associated smoothing benefits. Other studies in the area include Greasley and Oxley (1997), Linden (1995), and Demery and Duck (1992).

Impulse response and variance decomposition have been widely employed to observe cross-effects of shocks, evaluated on the basis of a pre-specified and tested multivariate model. In the last decade, one of the extensively studied areas is capital market links and interactions, owing to an increasingly integrating global financial market offering opportunities that never existed before or exhausted in the domestic market. Investigating capital market integration in the Pacific basin in the context of impulse response, Phylaktis (1999) studies specifically the speed of adjustment of real interest rates to long-run equilibrium following a shock in each of these markets. It is found that countries in the region are closely linked with world financial markets. Moreover, the association of these markets with Japan is stronger than that with the US. Tse *et al.* (1996) examine information transmission in three Eurodollar futures markets of Imm, Simex and Liffe. Employing impulse response analysis and variance decomposition which explores further the common factor in the cointegration system, it is found that the common factor is driven by the last trading market in the 24-hour trading sequence. Each of the markets impounds all the information and rides on the common stochastic trend during trading hours, and the three markets can be considered one continuously trading market. In a study of equity market linkages in ASEAN countries, Roca *et al.* (1998) use impulse response analysis and variance decomposition based on a VAR with error correction to investigate the extent and structure of price linkages among these markets. They find evidence of short-term linkages among all but the Indonesian market. But in the long-run, the linkages, if any, are weak. Specifically, the Malaysian market is the most influential, i.e. its shocks considerably contribute to the forecast variance in other markets; while the Singapore and Thailand markets have most strong interaction with other markets, i.e. shocks in the Singapore and Thailand markets account for a large proportion of forecast variance in other markets and, in the meantime, shocks in other markets attribute to a large amount of forecast variance in Singapore and Thailand markets. Finally, their results indicate that the Indonesian market is isolated and not linked with any other ASEAN market.

Impulse response has been widely applied to regional studies and real estate where the response to shocks from various sources is one of the major concerns. Baffoe-Bonnie (1998) analyses the effect of key macroeconomic variables on house prices and the stock of houses sold in the framework of VAR and impulse response analysis. The results suggest that macroeconomic variables produce cycles in housing prices and houses sold. Considerable amount of the forecast variance in the housing market can be attributed to shocks in the employment growth and mortgage rate at both national and regional levels. The study also reveals that the dynamic behaviour of housing prices and the number of houses sold vary substantially among different regions and at different time periods. Hort (2000) employs impulse response analysis based on the estimation of a VAR model of the after-tax mortgage rate, house prices and sales, to examine prices and turnover in the owner-occupied housing market. The empirical results in the paper support that the adjustment of house price expectations following a shock to demand is slow due to informational imperfections in the housing market.

There also exist asymmetries in buyers and sellers' responses such that sales are expected to respond prior to prices where buyers are assumed to respond prior to sellers. Tse and Webb (1999), concerned with the effectiveness of land tax and capital gain tax in curbing hoarding of land and speculation, evaluate the effects of property tax on housing in Hong Kong. Using an impulse response function, they demonstrate that the transaction tax has a dynamic negative impact on housing returns, as the imposition of capital gain tax impairs the liquidity of property transaction, lowers the rate of return on property investment, and reduces revenue from land sales. They also show that the capital gain tax is capitalised into housing prices.

Various other studies can be found in the areas of business cycles and monetary policy evaluation, real and nominal exchange rate behaviour and linkages, PPP, debt markets, employment, regions and sectors, in virtually any dynamic models involving the analysis of effect and cross-effect of shocks.

Notes

1 Pesaran and Shin (1998) and Microfit use the total variance of the orthogonal case in the denominator, so the components do not sum up to 100 per cent.
2 The restricted model involves deletion of the lagged variables, with the *t*-statistic of their coefficients being less than one, and re-estimation. Therefore, the implementation of the model is complicated and the model differs in every case, whereas the unrestricted model decides the lag length, then includes all lagged variables. So, the implementation and estimation are 'standard'.

Questions and problems

1   What is meant by persistence? How is persistence measured?
2   Compare persistence analysis and the test for unit roots.
3   Discuss the advantages of the procedure in this chapter to standardise the multivariate persistence measure and its rationale.
4   Describe impulse response analysis and its application in evaluating the impact of shocks and policy changes.
5   Why is orthogonalisation required in impulse response analysis?
6   What is meant by generalised impulse response analysis? Can generalised impulse response analysis avoid all the complications in orthogonalisation while achieving the same goal?
7   The contribution by the shock in each of the sources, expressed as a percentage of the total variance, sums to 100 per cent in this chapter. Discuss its rationale.
8   Collect data from various sources and test for persistence in the following time series:

(a) the spot foreign exchange rates of selected industrialised nations and developing economies *vis-à-vis* the US$, testing one individual time series each time,

    (b) GDP of selected countries, testing one individual time series each time,

    (c) nominal interests in selected countries, testing one individual time series each time.

What do you find of their characteristics?

9 Collect data from various sources and test for multivariate persistence in the following groups of time series:

    (a) the spot foreign exchange rates of selected industrialised nations *vis-à-vis* the US$,

    (b) the spot foreign exchange rates of selected developing economies *vis-à-vis* the US$,

    (c) GDP of selected countries,

    (d) nominal interests in selected countries.

What do you find of their characteristics?

10 Collect data from various sources and carry out (orthogonal) impulse response analysis in the following groups of time series:

    (a) sectoral output indices in the UK,

    (b) GDP of the UK, the US and Japan

    (c) stock market return indices of the UK, the US and Japan.

What do you find of their characteristics?

11 Collect data from various sources and carry out generalised impulse response analysis in the following groups of time series:

    (a) sectoral output indices in the UK,

    (b) GDPs of the UK, the US and Japan

    (c) stock market return indices of the UK, the US and Japan.

What do you find of their characteristics? Analyse the differences in your findings from (9) and (10).

References

Baffoe-Bonnie, J. (1998), The dynamic impact of macroeconomic aggregates on housing prices and stock of houses: a national and regional analysis, *Journal of Real Estate Finance and Economics*, **17**, 179–197.

Blanchard, O.J. and Quah, D. (1989), The dynamic effects of aggregate demand and supply disturbances, *American Economic Review*, **79**, 655–673.

Campbell, J.Y. and Mankiw, N.W. (1987a), Are output fluctuations transitory?, *Quarterly Journal of Economics*, **102**, 857–880.

Campbell, J.Y. and Mankiw, N.W. (1987b), Permanent and transitory components in macroeconomic fluctuations, *American Economic Review*, **77** (Papers and Proceedings), 111–117.

Cashin, P., Liang, H. and McDermott, C.J. (2000), How persistent are shocks to world commodity prices? *IMF Staff Papers*, **47**, 177–217.

Cochrane, J.H. (1988), How big is the random walk in GDP?, *Journal of Political Ecomony*, **96**, 893–920.

Dekker, A., Sen, K. and Young, M.R. (2001), Equity market linkages in the Asia Pacific Region: a comparison of the orthogonalised and generalised VAR approaches, *Global Finance Journal*, **12**, 1–33.

Demery, D. and Duck, N.W. (1992), Are economic fluctuations really persistent?, A reinterpretation of some international evidence, *Economic Journal*, **102**, 1094–1101.

Goerlich, P. (1992), Cochrane.src, *RATS, Estima*.

Greasley, D. and Oxley, L. (1997), Shock persistence and structural change, *Economic Record*, **73**, 348–362.

Hort, K. (2000), Prices and turnover in the market for owner-occupied homes, *Regional Science and Urban Economics*, **30**, 99–119.

Linden, M. (1995), Finnish GNP series 1954/I–1990/IV: small shock persistence or trend stationarity? Some evidence with variance ratio estimates, *Empirical Economics*, **20**, 333–349.

Mayadunne, G., Evans, M. and Inder, B. (1995), An empirical investigation of shock persistence in economic time series, *Economic Record*, **71**, 145–156.

Pesaran, M.H. and Shin, Y. (1998), Generalized impulse response analysis in linear multivariate models, *Economics Letters*, **58**, 17–29.

Pesaran, M.H., Pierse, R.G. and Lee, K.C., (1993), Persistence, cointegration and aggregation: a disaggregated analysis of output fluctuations in the US economy, *Journal of Econometrics*, **56**, 67–88.

Phylaktis, K. (1999), Capital market integration in the Pacific Basin region: an impulse response analysis, *Journal of International Money and Finance*, **18**, 267–287.

Priestley, M.B. (1996), *Spectral Analysis and Time Series* 9th edn (1st edn 1981), Academic Press, London.

Roca, E.D., Selvanathan, E.A. and Shepherd, W.F. (1998), Are the ASEAN equity markets interdependent?, *Asean Economic Bulletin*, **15**, 109–120.

Tse, R.Y.C. and Webb, J.R. (1999), Property tax and housing returns, *Review of Urban and Regional Development Studies*, **11**, 114–126.

Tse, Y., Lee, T.H. and Booth, G.G. (1996), The international transmission of information in eurodollar futures markets: a continuously trading market hypothesis, *Journal of International Money and Finance*, **15**, 447–465.

Van de Gucht, L.M., Dekimpe, M.G. and Kwok, C.C.Y. (1996), Persistence in foreign exchange rates, *Journal of International Money and Finance*, **15**, 191–220.

Wang, P.J. (2000), Shock persistence in property and related markets, *Journal of Property Research*, **17**, 1–21.

7 Modelling regime shifts
Markov switching models

Recent renewed interests in Markov chain processes and Markov switching models are largely fascinated by Hamilton (1989, 1994). While the major contributors with economic significance to the popularity of this family of models are the intensified studies in business cycles in the last two decades in the frontier of macro and monetary economics, and the proliferating use of mathematical tools in the exploitation of excess returns in a seemingly efficient while volatile financial market. The regime shift or state transition features of Markov switching, when applied properly, are able to illustrate and explain economic fluctuations around boom–recession or more complicated multi-phase cycles. In financial studies, the state transition process can be coupled with bull–bear market alternations, where regimes are less clearly defined but appear to have more practical relevance. However, estimation of Markov switching models may be technically difficult and the results achieved may be sensitive to the settings of the procedure. Probably, rather than producing a set of figures of immediate use, the approach helps improve our understanding about an economic process and its evolving mechanism constructively, as with many other economic and financial models.

7.1. Markov chains

A Markov chain is defined as a stochastic process $\{S_t, t = 0, 1, \ldots\}$ that takes finite or countable number of integer values denoted by i, j, and that the probability of any future value of S_{t+1} equals j, i.e., the conditional distribution of any future state S_{t+1}, given the past state $S_0, S_1, \ldots, S_{t-1}$ and the present state S_t, is only dependent on the present state and independent of the past states. That is:

$$P\left\{S_{t+1} = j \mid S_t = i_t, S_{t-1} = i_{t-1}, S_1 = i_1, S_0 = i_0\right\} = P\left\{S_{t+1} = j \mid S_t = i_t\right\} = p_{ij} \tag{7.1}$$

p_{ij} is the probability that the state will next be j when the immediate preceding state is i, and can be called the transition probability from i into j. Suppose there

are N states, then all the transitions can be expressed in a transition matrix:

$$
\mathbf{P} = \begin{bmatrix}
p_{11} & p_{12} & \cdot & \cdot & p_{1N} \\
p_{21} & p_{22} & \cdot & \cdot & p_{2N} \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot & \cdot \\
p_{N1} & p_{N2} & \cdot & \cdot & p_{NN}
\end{bmatrix}
\tag{7.2}
$$

The probability is non-negative and the process must transit into some state, including the current state itself, so that:

$$
\sum_{j=1}^{N} p_{ij} = 1, \quad i = 1, 2, \ldots, N
\tag{7.3}
$$

Above are one-step transition probabilities. It is natural for us to extend the one-step case and consider n-step transitions that are clearly functions and results of several one-step transitions. For example, a two-step transition $P\left\{ S_{t+2} = j \mid S_t = i \right\}$ probability is the summation of the probabilities of transitions from state i into all the states, then from all the states into state j:

$$
\sum_{k=1}^{N} P\left\{ S_{t+2} = j \mid S_{t+1} = k \right\} P\left\{ S_{t+1} = k \mid S_t = i \right\}
$$

More generally, define the n-step transition probability as:

$$
P\left\{ S_{t+n} = j \mid S_t = i \right\} = p_{ij}^n
\tag{7.4}
$$

A formula called the Chapman–Kolmogorov equation holds for calculating multi-step transition probabilities:

$$
p_{ij}^{m+n} = \sum_{k=1}^{N} p_{ik}^n p_{kj}^m, \quad i, j = 1, 2, \ldots N
\tag{7.5}
$$

7.2. Estimation

The estimation of a Markov chain process or Markov switching model is achieved, naturally, by considering the joint conditional probability of each of future states, as a function of the joint conditional probabilities of current states and the transition probabilities. This procedure is called filtering: the conditional probabilities of current states are input, passing through or being filtered by the system of dynamic transformation that is the transition probability matrix, to produce the conditional probabilities of future states as output. The conditional likelihood function can be obtained in the meantime, and the parameter can be estimated accordingly.

Suppose there is a simply two-state Markov chain process:

$$y_t = \mu_1 S_1 + \mu_2 S_2 + \varepsilon_t \tag{7.6}$$

where $S_1 = 1$ when in state 1 and 0 otherwise, $S_2 = 1$ when in state 2 and 0 otherwise, and ε_t is a white noise residual. We are interested to know how the joint probability of y_t and S_t transits over time. This can be achieved in two major steps. The first is to have an estimate of the conditional probability $P(S_t = s_t | y_{t-1})$, i.e., the probability of being in state s_t, based on information available at time $t-1$. According to the transition probability and property, that is straightforward. The second is to consider the joint probability density distribution of y_t and S_t, so the probability of being in state s_t is updated to $P(S_t = s_t | y_t)$, using information available at time t. The procedure is as follows:

(i) estimating the probability of being in state s_t, conditional on information at $t-1$:

$$P(S_t = s_t | y_{t-1}) = P(S_t = s_t | S_{t-1} = s_{t-1}) \times P(S_{t-1} = s_{t-1} | y_{t-1})$$

(ii) (a) calculating the joint density distribution of y_t and S_t:

$$f(y_t, S_t = s_t | y_{t-1})$$
$$= f(y_t | S_t = s_t, y_{t-1}) \times P(S_t = s_t | y_{t-1})$$
$$= f(y_t | S_t = s_t, y_{t-1}) \times P(S_t = s_t | S_{t-1} = s_{t-1}) \times P(S_{t-1} = s_{t-1} | y_{t-1}) \tag{7.7}$$

(b) calculating the density distribution of y_t:

$$f(y_t | y_{t-1}) = \sum_{s_t=1}^{2} f(y_t, S_t = s_t | y_{t-1}) \tag{7.8}$$

(c) calculating the following:

$$P(S_t = s_t | y_t) = \frac{f(y_t, S_t = s_t | y_{t-1})}{f(y_t | y_{t-1})} \tag{7.9}$$

that is the updated joint probability of y_t and S_t.

Consider now a general N-state Markov chain process y_t that has autoregression of order r in its residual ε_t and is also the function of the exogenous variable x_t and its lags. This is the typical dynamic process of autoregression, frequently encountered in contemporary empirical economics and finance, if there is only one state. When variable y_t in a Markov chain process has autoregression of order r,

the joint conditional probability of the current state and r previous states, based on information set including all its lags up to r periods before period 0, i.e.:

$$P(S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-1} = s_{t-r} | \Omega_{t-1}) \tag{7.10}$$

should be considered, where $\Omega_{t-1} = (y_{t-1}, y_{t-2}, \ldots y_{-r}, x_{t-1}, x_{t-2}, \ldots x_{-r})$ is the information set available at time $t-1$. The filtering procedure, which is to update the joint conditional probability of equation (7.6) from the previous joint conditional probability, is as follows:

(1) calculating the joint density distribution of y_t and S_t:

$$f(y_t, S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1} | \Omega_{t-1})$$

$$= f(y_t | S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1}, \Omega_{t-1})$$

$$\times P(S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1} | \Omega_{t-1})$$

$$= f(y_t | S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1}, \Omega_{t-1})$$

$$\times P(S_t = s_t | S_{t-1} = s_{t-1}) \times P(S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1} | \Omega_{t-1}) \tag{7.11}$$

(2) calculating the density distribution of y_t:

$$f(y_t | \Omega_{t-1})$$

$$= \sum_{s_t=1}^{N} \sum_{s_{t-1}=1}^{N} \cdots \sum_{s_{t-r}=1}^{N} f(y_t, S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1} | \Omega_{t-1}) \tag{7.12}$$

(3) calculating the following that, unlike the non-serial correlation residual case, is not yet the output of the filter:

$$P(S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1} | \Omega_t)$$

$$= \frac{f(y_t, S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1} | \Omega_{t-1})}{f(y_t | \Omega_{t-1})} \tag{7.13}$$

(4) the output of the filter is then the summation over the states at lag r:

$$P(S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r} = s_{t-r} | \Omega_t)$$

$$= \sum_{s_{t-r-1}=1}^{N} P(S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r-1} = s_{t-r-1} | \Omega_t) \tag{7.14}$$

During the above course the probability of the states at time t, based on currently available information, is obtained:

$$P(S_t = s_t \mid \Omega_t) = \sum_{s_t=1}^{N} \sum_{s_{t-1}=1}^{N} \cdots \sum_{s_{t-r}=1}^{N} P(S_t = s_t, S_{t-1} = s_{t-1}, \ldots S_{t-r} = s_{t-r} \mid \Omega_t)$$

(7.15)

The log likelihood function is also derived:

$$L(\theta) = \sum_{t=1}^{T} f(y_t \mid \Omega_{t-1}; \theta)$$

(7.16)

where θ represents the vector of parameters. There are few techniques that are singled out for estimating the log likelihood function, such as the Gibbs sampling and the EM algorithm, but maximum likelihood remains a useful, convenient and largely appropriate method in practice. Maximising equation (7.16) leads to derivation of the estimates with regard to parameters and states.

Using the simple instance of the two-state Markov chain process of equation (7.6) and assuming a normally distributed residual, we write down its maximum likelihood function explicitly, that can be routinely extended to more complicated cases, as follows:

$$L(\theta) = \sum_{t=1}^{T} f(y_t \mid y_{t-1}; \theta)$$

$$= \sum_{t=1}^{T} \sum_{s_t=1}^{2} f(y_t \mid S_t = s_t, y_{t-1}; \theta) \times P(S_t = s_t \mid y_{t-1})$$

$$= \sum_{t=1}^{T} \sum_{s_t=1}^{2} \sum_{s_{t-1}=1}^{2} \{ f(y_t \mid S_t = s_t, y_{t-1}; \theta) \times P(S_t = s_t \mid S_{t-1} = s_{t-1})$$

$$\times P(S_{t-1} = s_{t-1} \mid y_{t-1}) \}$$

$$= \sum_{t=1}^{T} \left\{ \frac{1}{\sqrt{2\pi} \sigma_\varepsilon} \exp\left(\frac{-(y_t - \mu_1)^2}{2\sigma_\varepsilon^2} \right) \times \left[p_{11} \times P_{tL}(1) + p_{21} \times P_{tL}(2) \right] \right.$$

$$\left. + \frac{1}{\sqrt{2\pi} \sigma_\varepsilon} \exp\left(\frac{-(y_t - \mu_2)^2}{2\sigma_\varepsilon^2} \right) \times \left[p_{21} \times P_{tL}(1) + p_{22} \times P_{tL}(2) \right] \right\}$$

(7.17)

where $P_{tL}(1) = P(S_{t-1} = 1 \mid y_{t-1})$ and $P_{tL}(2) = P(S_{t-1} = 2 \mid y_{t-1})$ for simplicity.

7.3. Smoothing

Similar to the case of the Kalman filter to be introduced in Chapter 9, the states at time t have been estimated based on the information set at t in the

above procedure. It may be of interest to review the states at a later time when more information is available, or infer the states using the whole information set up to the last observation at time T. An inference made about the present states using future information is called smoothing, with the inference made with the whole information set being full smoothing, or simply smoothing. Smoothing may be of no use to problems such as real time control in cybernetics, but it provides more desirable results when an insightful understanding of the process is the major concern; for example, in the economic science for revealing the working mechanism of dynamic economic systems and shaping future policies.

Smoothing is to revise $P(S_t = s_t | \Omega_t)$, the probability of the states at time t based on currently available information, to $P(S_t = s_t | \Omega_T)$, the probability of the states at time t based on the whole information set. Put simply, it replaces Ω_t by Ω_T in the probability. Smoothing involves two steps when there is no lag in y_t, and three steps and one approximation when there are lags in y_t.

(1) Calculating (to save space, $S_t = s_t$ has been simplified as S_t):

$$P(S_{t-r}, \ldots S_t, S_{t+1} | \Omega_T)$$

$$= P(S_{t-r+1}, \ldots S_t, S_{t+1} | \Omega_T) \times P(S_{t-r} | S_{t-r+1}, \ldots S_t, S_{t+1}, \Omega_T)$$

$$= P(S_{t-r+1}, \ldots S_t, S_{t+1} | \Omega_T) \times P(S_{t-r} | S_{t-r+1}, \ldots S_t, S_{t+1}, \Omega_t)$$

$$= \frac{P(S_{t-r+1}, \ldots S_{t+1} | \Omega_T) \times P(S_{t-r}, \ldots S_{t+1}, \Omega_t)}{P(S_{t-r+1}, \ldots S_t, S_{t+1} | \Omega_t)}$$

$$= \frac{P(S_{t-r+1}, \ldots S_{t+1} | \Omega_T) \times P(S_{t-r}, \ldots S_t, \Omega_t) \times P(S_{t+1} | S_t)}{P(S_{t-r+1}, \ldots S_t, S_{t+1} | \Omega_t)} \tag{7.18}$$

The second equality involving $P(S_{t-r} | S_{t-r+1}, \ldots S_t, S_{t+1}, \Omega_T) = P(S_{t-r} | S_{t-r+1}, \ldots S_t, S_{t+1}, \Omega_t)$ is exact only if:

$$f(y_{t+1}, \Omega_{Tt} | S_{t-r}, S_{t-r+1} \ldots S_t, S_{t+1}, \Omega_t)$$
$$= f(y_{t+1}, \Omega_{Tt} | S_{t-r+1}, \ldots S_t, S_{t+1}, \Omega_t) \tag{7.19}$$

holds. It is because, define $\Omega_{Tt} = \Omega_T - \Omega_t$, it follows:

$$P(S_{t-r} | S_{t-r+1} \ldots S_t, S_{t+1}, \Omega_T) = P(S_{t-r} | S_{t-r+1}, \ldots S_t, S_{t+1}, \Omega_t, \Omega_{Tt})$$

$$= \frac{f(y_{t+1}, S_{t-r}, \Omega_{Tt} | S_{t-r+1} \ldots S_t, S_{t+1}, \Omega_t)}{f(y_{t+1}, \Omega_{Tt} | S_{t-r+1} \ldots S_t, S_{t+1}, \Omega_t)}$$

$$= \frac{f(y_{t+1}, \Omega_{Tt} | S_{t-r} \ldots S_t, S_{t+1}, \Omega_t) \times P(S_{t-r} | S_{t-r+1} \ldots S_t, S_{t+1}, \Omega_t)}{f(y_{t+1}, \Omega_{Tt} | S_{t-r+1} \ldots S_t, S_{t+1}, \Omega_t)}$$

$$\tag{7.20}$$

(2) Summing up over $S_{t+1} = 1, 2, \ldots N$:

$$P(S_{t-r}, \ldots S_t \,|\, \Omega_T) = \sum_{s_{t+1}=1}^{N} P(S_{t-r}, \ldots S_t, S_{t+1} \,|\, \Omega_T) \qquad (7.21)$$

Equation (7.21) is already the smoothed states when there is no serial correlation in the residual or there is no lagged y_t involved. When there are lags, smoothing is similar to equation (7.15), finally achieved through the following summation:

$$P(S_t \,|\, \Omega_T) = \sum_{s_t=1}^{N} \sum_{s_{t-1}=1}^{N} \cdots \sum_{s_{t-r}=1}^{N} P(S_t, S_{t-1}, \ldots S_{t-r} \,|\, \Omega_T) \qquad (7.22)$$

7.4. Time-varying transition probabilities

It is natural to extend the above analysis to allow the Markov chain model additional flexibility, by introducing time-varying transition probabilities. Let us define the time-varying transition probability as follows:

$$P\{S_{t+1} = j \,|\, S_t = i_t, \, |\, \Omega_{t+1}\} = p_{ij}(t+1) \qquad (7.1')$$

Then the transition probability matrix is:

$$\mathbf{P(t)} = \begin{bmatrix} p_{11}(t) & p_{12}(t) & \cdots & p_{1N}(t) \\ p_{21}(t) & p_{22}(t) & \cdots & p_{2N}(t) \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ p_{N1}(t) & p_{N2}(t) & \cdots & p_{NN}(t) \end{bmatrix} \qquad (7.2')$$

The choice of types of time-varying transition probabilities is an empirical issue, though those used in binary choice models in the form of probit and logit are logically adopted, with the similar rationale argued for the probit and logit model. In addition, there is the exponential function and the cumulative normal distribution function. The exponential function and the cumulative normal distribution function are symmetric, with the mirror image on the vertical axis, so any departure from the mean value will increase the probability. While a logic function is asymmetric with a positive departure and a negative departure from the mean value having opposite effects. These time-varying functions are also similar to what are widely used in smoothing transition models. The use of time-varying transition probabilities has an additional advantage, that is, such specifications limit the value of the probability in the range of $[0, 1]$ at the same time, or indeed, in any desirable ranges. This prevents unreasonable outcome from occurring in the execution of a programme. Even if the transition probability is not time-varying, using some functional forms to set the range of the probability is always helpful.

The logit function of transition probabilities is:

$$p_{ij}(t) = \frac{1}{1 + \exp\left\{-\Omega_t \boldsymbol{\beta}'_{ij}\right\}} \tag{7.23}$$

where $\boldsymbol{\beta}'_{ij}$ is a vector of coefficients on the set of dependent and exogenous variables. $\exp\left\{-\Omega_t \boldsymbol{\beta}'_{ij}\right\}$ can change from 0 to ∞, containing the probability in the range of $[0, 1]$. In a simple example, when $-\Omega_t \boldsymbol{\beta}'_{ij} = \omega_{ij0} - \gamma_{ij} y_{t-1}$, equation (7.23) becomes:

$$p_{ij}(t) = \frac{1}{1 + \exp\left\{\omega_{ij0} - \gamma_{ij} y_{t-1}\right\}}$$

It has a mean value of 0.5 when $y_t = \omega_{ij}/\gamma_{ij}$ and will increase when $y_t > \omega_{ij}/\gamma_{ij}$ with $p_{ij}(t) \underset{y_{t-1} \to \infty}{\to} 0$ decrease when $y_t < \omega_{ij}/\gamma_{ij}$ with $p_{ij}(t) \underset{y_{t-1} \to -\infty}{\to} 1$, provided γ_{ij} is positive. A cumulative normal distribution has the similar pattern.

An exponential type transition probability is specified as follows:

$$p_{ij}(t) = 1 - \exp\{-(\Omega_t \boldsymbol{\beta}'_{ij})^2\} \tag{7.24}$$

$\exp\{-(\Omega_t \boldsymbol{\beta}'_{ij})^2\}$ can change from 1 to 0, limiting the probability in the range of $[0, 1]$. Using the same example of $-\Omega_t \boldsymbol{\beta}'_{ij} = \omega_{ij0} - \gamma_{ij} y_{t-1}$, equation (7.24) becomes:

$$p_{ij}(t) = 1 - \exp\{-(\omega_{ij0} - \gamma_{ij} y_{t-1})^2\}$$

It has the maximum value of unity when $y_t = \omega_{ij}/\gamma_{ij}$, and will decrease when y_t departs from ω_{ij}/γ_{ij}, no matter whether $y_t - (\omega_{ij}/\gamma_{ij})$ is positive or negative.

The above two specifications have direct economic meanings and implications, e.g., symmetric responses related only to the distance of departure from a central point or the equilibrium, no matter what is the direction or the sign of departure; and asymmetric effects where both the distance and the sign are relevant. If the purpose is to restrict the value of the probability only, then many simpler and more straightforward specifications, such as the one used in Example 7.1 in section 7.5, can perform satisfactorily.

7.5. Examples and cases

Example 7.1

We use the Markov chain model to illustrate regime shifts in business cycle conditions in UK GDP data at the factor price running from the first quarter of 1964 to the fourth quarter of 1999.

Table 7.1 Estimation of UK GDP with a two-regime Markov switching model: 64Q1–99Q4

μ_1	$0.7491e^{-2}$***	$(0.1281e^{-2})$
μ_2	$-0.1517e^{-1}$***	$(0.5446e^{-2})$
σ_1	$0.8591e^{-4}$***	$(0.7592e^{-5})$
σ_2	$0.6626e^{-4}$	$(0.4354e^{-4})$
ω_{11}[a]	3.2153***	(1.1569)
ω_{22}[b]	0.7245	(1.0248)

***significant at the 1 per cent level. Standard errors in parentheses.

[a]The parameter from using a simple function, $p = e^{\omega}/(1 + e^{\omega})$, to impose restrictions on the range of the probability. p_{11}, the transition probability of staying in normal periods, is 0.9613, according to the function.

[b]Equivalent to a p_{22}, the transition probability of remaining in a recession, of 0.6736.

The model has two means for recessions and normal times respectively. The residual follows an autoregressive process of order 1 and has different volatility or variance in the two regimes. Let y_t be the logarithm of GDP, S_1 be state for normal times, and S_2 be state for recessions:

$$\Delta y_t = \mu + \mu_2 S_2 + \rho \Delta y_{t-1} + \omega_t$$
$$\omega_t \sim (0, S_1\sigma_1^2 + S_2\sigma_2^2)$$

(7.25)

With this specification, the growth rate is μ in normal times, and $\mu + \mu_2$ in recessions. While the variance is σ_1^2 in normal periods and σ_2^2 in recessions. We adopt equation (7.23) to restrict the transition probability in the range of [0, 1], though the transition probability is not time-varying. The results from estimating the model are reported in Table 7.1.

It has been found that UK GDP growth is about 0.7 per cent per quarter (μ_1), translating into an annual growth rate of 3 per cent, during normal times in the estimation period. In recessions, the growth rate is a negative 0.7 per cent ($\mu_1 + \mu_2$), or a negative 3 per cent per annum. The transition probability of staying in normal periods, or from normal to normal is 0.9613, being calculated from $\omega_{11,0}$ and using a simple function, $p = e^{\omega}/(1 + e^{\omega})$, to impose restrictions on the range of the probability. With similar transformation, the transition probability of remaining in a recession is 0.6736. This transition probability is, however, statistically insignificant and therefore unreliable. One of the reasons is that the duration of recessions is relatively short, so the probability of staying in the recession varies, especially when the economy is nearing the end of a recession. The duration

Continued

of being in normal times is:

$$\frac{1}{1-p_{11}} = \frac{1}{1-0.9613} \approx 26 \text{ quarters or } 6.5 \text{ years}$$

The duration of an average recession is:

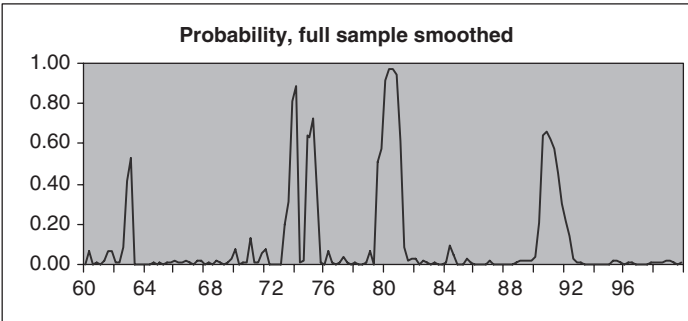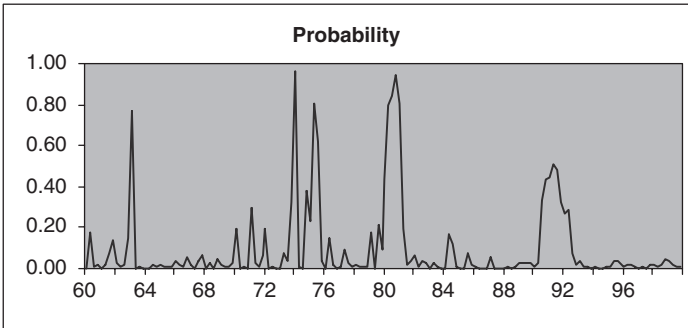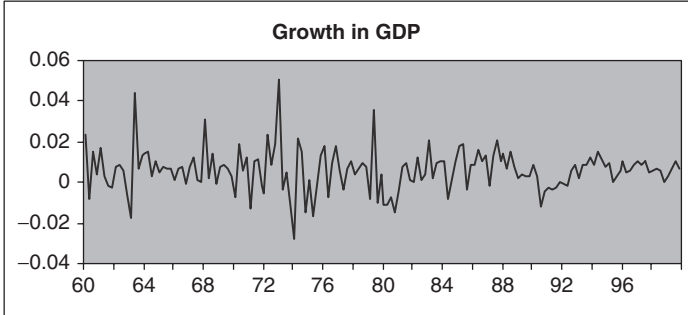$$\frac{1}{1-p_{22}} = \frac{1}{1-0.6736} \approx 3 \text{ quarters}$$

Figure 7.1 Growth in UK GDP.

As observed before, the errors associated with p_{22} are large so the duration of recessions could well deviate from three quarters by a large margin. The two regimes also have different volatility. In normal times, the standard deviation of the residual is $0.8591e^{-4}$ (σ_1), or about 0.009 per cent per quarter, being statistically significant at the 1 per cent level. While the standard deviation seems smaller in recessions with σ_2 being $0.6626e^{-4}$, it does not suggest lower volatility as the statistic is statistically insignificant. Since recession periods are relatively short with much fewer observations being available, this statistic is unreliable. We can see from Table 7.1 that the standard error of it is $0.4354e^{-4}$. So the standard deviation of the residual can be very large as well as very small. This does cast more uncertainty in recessions.

The business cycle regime characteristics of UK GDP are exhibited in Figure 7.1. Notice the probability of being in one of the states is time-varying, regardless whether the transition probabilities are constant or not. Panel (a) in Figure 7.1 is the growth rate of UK GDP between the first quarter in 1964 and the fourth quarter in 1999. Panel (b) shows the probability of being in the state of recession without smoothing, and Panel (c) is the full sample smoothed probability for the same state. As in most empirical studies, there is only very small difference between the two representations of probability.

Example 7.2

Oil price volatility has long been considered a factor influencing the state of business cycles and, in particular, plunging the economy into recessions when there is a sharp increase in the oil price, or an oil price crisis. Therefore, the oil price is frequently used as a variable of impact in time-varying transition probabilities. One of the examples is a study by Raymond and Rich (1997) entitled 'Oil and the macroeconomy: a Markov state-switching approach'. Their modelling of time-varying transition probabilities follows Filardo (1994); and the treatment of the oil price series follows Hamilton (1996), having considered the asymmetric effects of oil price changes on business cycles. The net oil price increase variable proposed by Hamilton (1996) is equal to the percentage change in the current real oil price above the maximum of the previous four quarters if positive and zero otherwise. Bearing this characteristic in mind, their mean equation is:

$$\Delta y_t = \alpha_0 + \alpha_1 S_t + \sum_{i=1}^{n} \beta_i o_{t-i}^+ + \varepsilon_t, \quad \alpha_1 < 0$$

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

(7.26)

Continued

where $S_t = 0$ is the state for the normal period or with higher growth rate, $S_t = 1$ the state for recessions, and o_t^+ is the net oil price increase variable explained in the above. There is no lagged real GDP growth entering the mean equation. The specification does not distinguish the volatility or variance of the residual between the higher growth period and recession. The time-varying transition probabilities are designed as follows:

$$P\{S_t = 0 \,|\, S_{t-1} = 0, o_{t-1}^+, o_{t-2}^+, \ldots\} = q_t = \Phi\left(\delta_0 + \sum_{i=1}^{d} \delta_i o_{t-i}^+\right)$$

$$P\{S_t = 1 \,|\, S_{t-1} = 1, o_{t-1}^+, o_{t-2}^+, \ldots\} = p_t = \Phi\left(\gamma_0 + \sum_{i=1}^{d} \gamma_i o_{t-i}^+\right)$$

(7.27)

where $\Phi(\cdot)$ is the cumulative normal distribution function with the same purpose as in Example 1 to limit the range of the transition probability between 0 and 1.

The data sample period in the study is from the first quarter of 1951 to the third quarter in 1995 for both US real GDP and the real price of oil. The empirical results are summarised in Table 7.2, where the quarterly growth rate has been multiplied by 100.

The unrestricted model, where the net oil price increase variable entering both the mean equation for real GDP growth and the transition probability, has achieved the highest log likelihood function value. Comparing the two restricted versions with the general model of no restriction by the statistic of the likelihood ratio test, however, it is found that the time-varying transition probability model is of no difference from a constant transition probability model. That is the validity of the restriction cannot be rejected at any conventional statistical significance levels, with a log likelihood ratio being LR = 0.894. Nevertheless, the oil variable plays a role in the mean equation and the restriction is rejected by a log likelihood ratio test statistic of 12.142. The above analysis suggests that the net oil price increase variable has a negative impact on the growth of real GDP but provides little valid information about future switches between the two regimes and their timing. Indeed, none of the coefficients for lagged net oil price increases are statistically significant in the transition probabilities with the general model; and only the coefficient for the net oil price increase variable at lag 4 in the transition probability of remaining in the normal time (δ_4) is significant at the 5 per cent level with the model where restrictions are imposed on the mean equation. But, as indicated earlier, that restrictions on the coefficients in the mean equation are rejected, so estimates obtained with that model are questionable. Besides, none of the parameters in the transition probability of remaining in recession, either the constant or the coefficients for lagged

Table 7.2 Estimation of US real GDP with a time-varying transition probability Markov switching model: 51Q1–95Q3

	Restricted: oil has no effect on transition probabilities	*Restricted: oil has no effect in the mean equation*	*No restrictions: the general model*
α_0	1.066***	0.929***	1.018***
	(0.097)	(0.076)	(0.081)
$\alpha_0 + \alpha_1$	−0.068	−0.593**	−0.081
	(0.310)	(0.294)	(0.341)
β_1	−0.031***	–	−0.026**
	(0.012)		(0.012)
β_2	−0.013	–	−0.008
	(0.012)		(0.013)
β_3	−0.027**	–	−0.032***
	(0.012)		(0.013)
β_4	−0.046***	–	−0.021
	(0.011)		(0.014)
δ_0	1.484***	1.866***	1.750***
	(0.375)	(0.322)	(0.361)
δ_3	–	−0.053	−0.044
		(0.069)	(0.051)
δ_4	–	−0.154**	−0.139
		(0.073)	(0.090)
γ_0	0.334	1.012	0.779
	(0.387)	(1.008)	(0.704)
γ_3	–	0.918	0.948
		(0.846)	(1.043)
γ_4	–	−0.485	−0.502
		(0.344)	(0.410)
σ_ε	0.714***	0.753***	0.732***
	(0.050)	(0.047)	(0.048)
Log likelihood	−209.320	−214.944	−208.873

**significant at the 5 per cent level; ***significant at the 1 per cent level. Standard errors in parentheses.

oil price increases, are statistically significant. This is consistent with the findings in Example 7.1.

From applying equation (7.27) and the estimates in Table 7.2, q, the average transition probability of remaining in the normal period, is 0.931; and p, the average transition probability of remaining in recession, is 0.631. The average duration of being in normal times is $1/(1-q) = 1/(1-0.931) \approx 14.5$ quarters or slightly more than 3.5 years. The duration

Continued

of an average recession is $1/(1-p) = 1/(1-0.631) \approx 2.7$ quarters. These durations, especially the duration of normal periods, are relatively shorter than those in Example 7.1 with the UK case. The difference may suggest that the UK economy has a longer duration of normal periods but suffers more severely in recessions, or arise from the sensitivity of the parameters to estimation procedures and data sets.

7.6. Empirical literature

Markov switching approaches have attracted much attention in financial and economic modelling in recent years, due to business cycle characteristics high-lighted in macroeconomics and monetary economics, and a changing business and investment environment featured by bull–bear market alternations in financial studies. Collectively, these cyclical movements can be termed as regime shifts, common to most modern market economies. As the Markov switching model clearly defines two or more states or regimes, it can reveal the dynamic process of the variables of concern vividly and provide the researcher and policy maker with a clue of how these variables have evolved in the past and how they may change in the future. Nevertheless, the implementation and execution of a Markov switching model, though not complicated, may be technically difficult as it is rather sensitive to the choice of initial values, other settings such as the lag length, and even the data sample.

Stock market behaviour is one of the areas to which Markov switching has been widely applied. In a paper entitled 'Identifying bull and bear markets in stock returns', Maheu and McCurdy (2000) use a Markov switching model to classify returns into a high-return stable state and a low-return volatile state. They call the two states bull and bear markets respectively. Using 160 years' US monthly data, they find that bull markets have a declining hazard function although the best market gains come at the start of a bull market. Volatility increases with duration in bear markets. Driffill and Sola (1998) investigate whether there is an intrinsic bubble in stock prices so that stock prices deviate from the values predicted by the present value model or deviate from the fundamental relationship between income and value. They claim that a Markov switching model is a more appropriate representation of dividends. Allowing for dividends to switch between regimes they show that stock prices can be better explained than by the bubble hypothesis. When both the bubble and the regime switching in the dividend process are considered, the incremental explanatory contribution of the bubble is low. Assoe (1998) examines regime switching in nine emerging stock market returns. The author claims that changes in government policies and capital market reforms may lead to changes in return generating processes of capital markets. The results show strong evidence of regime switching behaviour in emerging

stock market returns with regard to volatility which foreign investors concern most. Other research includes Dewachter and Veestraeten (1998) on jumps in asset prices which are modelled as a Markov switching process in the tradition of event studies; Scheicher's (1999) investigation into the stock index of the Vienna Stock Exchange with daily data from 1986 to 1992, adopting Markov switching and GARCH alternatives; and So *et al.* (1998) who examine the S&P500 weekly return data with the Markov switching approach to modelling stochastic volatility, and identified high, medium, and low volatility states associated with the return data.

The business and investment environment can be reasonably characterised by switching between different regimes as well. In this regard, Asea and Blomberg (1998) investigate the lending behaviour of banks over lending cycles, using the Markov switching model with a panel data set consisting of approximately two million commercial and industrial loans granted by 580 banks between 1977 and 1993. They demonstrate that banks change their lending standards from tightness to laxity systematically over the cycle. Town (1992), based on the well observed phenomenon that mergers take place in waves, fits the merger data into a Markov switching model with shifts between two states of high and low levels of activity and claims improvements over ARIMA models.

The changing pattern of interest rates is indicative of business cycle conditions and could be subject to regime shifts itself. To investigate how real interest rates shift, Bekdache (1999) adopts a time varying parameter model with Markov switching conditional heteroscedasticity to capture two sources of shifts in real interest rates: shifts in the coefficients and shifts in the variance. The former relates the *ex ante* real rate to the nominal rate, the inflation rate and a supply shock variable, and the latter is unconditional shifts in the variance of the stochastic process. The results prefer a time varying parameter model to Markov switching with limited states. Dewachter (1996) studies interest rate volatility by examining both regime shifts in the variance and links between volatility and levels of the interest rate. While regime shifts are found in the variance, the contribution of volatility-level links cannot be ignored. The above findings suggest that univariate or single element regime shifts in interest rate modelling fail to fully characterise interest rate dynamics.

Probably the majority of applied research is in the area of business cycles where recent studies are still burgeoning. In addition to classifying the economy into two states of booms and recessions, Kim and Nelson (1999) further investigate whether there has been a structural break in post-war US real GDP growth towards stabilisation. They employ a Bayesian approach to identifying a structural break at an unknown change-point in a Markov-switching model. Their empirical results suggest a break in GDP growth toward stabilisation at the first quarter of 1984, and a narrowing gap between growth rates during recessions and booms. Filardo and Gordon (1998) specify a time varying transition probability model where the information contained in leading indicator is used to forecast transition probabilities and, in turn, to calculate expected business cycle durations.

Both studies employ Gibbs sampling techniques. Other research in the category covers Diebold *et al.* (1993), Filardo (1994), Ghysels (1994), Luginbuhl and de Vos (1999), Kim and Yoo (1996), and Raymond and Rich (1997) as illustrated in Example 7.2.

It should be noted that the empirical application of Markov switching models is not always superior to an alternative or simple model, and is not without deficiencies. Aware of these problems, Boldin (1996) explores the robustness of Hamilton's (1989) two-regime Markov switching model framework. Applying Hamilton's exact specification to a revised version of real GNP, the author finds that parameter estimates are similar to those reported by Hamilton only when the author uses the same sample period (1952–1984) and a particular set of initial values for the maximum likelihood procedure. Two other local maximums exist that have higher likelihood values, and neither correspond to the conventional recession–expansion dichotomy. When the sample period is extended, there is no longer a local maximum near the parameter set reported by Hamilton. Exploring the model and data further, the author rejects cross-regime restrictions of Hamilton specification, but also finds that relaxing these restrictions increases the number of local maximums. In a study on the prediction of US business cycle regimes, Birchenhall *et al.* (1999) compare the use of logistic classification methods and Markov switching specifications for the identification and prediction of post-war US business cycle regimes as defined by the NBER reference turning point dates. They examine the performance of logistic procedures in reproducing the NBER regime classifications and in predicting one and three months ahead growth rates using leading indicator variables. They show that the logistic classification model provides substantially more accurate business cycle regime predictions than the Markov switching model. Nevertheless, as said at the beginning of this chapter, one of the major contributions the Markov switching approach made is probably to help improve our understanding about an economic process. This may partly explain its contemporary popularity. In addition to above discussed empirical literature, a variety of applications can be further found in foreign exchange rates, bond yields, inflation, and so on.

Questions and problems

1   Describe the state and the state transition probability in a Markov chain.
2   What is the Chapman–Kolmogorov equation for calculating multi-step transition probabilities?
3   Cite examples of economic and financial variables which can be shown as a Markov process.
4   What is smoothing is the estimation of a Markov process? Why is smoothing required?
5   Discuss the advantages of adopting time-varying transition probabilities in the Markov process.

6 Collect data from various sources, and estimate a two-state constant transition probability model in the following time series (using RATS, GAUSS or other packages):

  (a) industrial production of selected countries,
  (b) CPI of the G7,
  (c) GDP of the US, Argentina, France, Algeria, and India.

7 Estimate a two-state time-varying transition probability model in the above time series.

References

Asea, P.K. and Blomberg, B. (1998), Lending cycles, *Journal of Econometrics*, **83**(1–2), 89–128.

Assoe, K.G. (1998), Regime-switching in emerging stock market returns, *Multinational Finance Journal*, **2**(2), 101–132.

Bekdache, B. (1999), The time-varying behaviour of real interest rates: a re-evaluation of the recent evidence, *Journal of Applied Econometrics*, **14**(2), 171–190.

Birchenhall, C.R., Jessen, H., Osborn, D.R. and Simpson, P. (1999), Predicting U.S. business-cycle regimes, *Journal of Business and Economic Statistics*, **17**(3), 313–323.

Boldin, M.D. (1996), A check on the robustness of Hamilton's Markov switching model approach to the economic analysis of the business cycle, *Studies in Nonlinear Dynamics and Econometrics*, **1**(1), 35–46.

Dewachter, H. (1996), Modelling interest rate volatility: regime shifts and level links, *Weltwirtschaftliches Archiv*, **132**(2), 236–258.

Dewachter, H. and Veestraeten, D. (1998), Expectation revisions and jumps in asset prices, *Economics Letters*, **59**(3), 367–372.

Diebold, F.X., Lee, J.H. and Weinbach, G.C. (1993), Regime switching with time varying transition probabilities, *Federal Reserve Bank of Philadelphia Research Working Paper*, 93–12.

Driffill, J. and Sola, M. (1998), Intrinsic bubbles and regime-switching, *Journal of Monetary Economics*, **42**(2), 357–373.

Filardo, A.J. (1994), Business cycle phases and their transitional dynamics, *Journal of Business and Economic Statistics*, **12**(3), 299–308.

Filardo, A.J. and Gordon, S.F. (1998), Business cycle durations, *Journal of Econometrics*, **85**(1), 99–123.

Ghysels, E. (1994), On the periodic structure of the business cycle, *Journal of Business and Economic Statistics*, **12**(3), 289–298.

Hamilton, J.D. (1989), A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**(2), 357–384.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Hamilton, J.D. (1996), This is what happened to the oil price-macroeconomy relationship, *Journal of Monetary Economics*, **38**, 215–220.

Kim, C.J. and Nelson, C.R. (1999), Has the U.S. economy become more stable? a Bayesian approach based on a Markov-switching model of the business cycle, *Review of Economics and Statistics*, **81**(4), 608–616.

Kim, M.J. and Yoo, J.S. (1996), A Markov switching factor model of coincident and leading indicators, *Journal of Economic Rese*arch, **1**(2), 253–72.

Luginbuhl, R. and de Vos, A. (1999), Bayesian analysis of an unobserved-component time series model of GDP with Markov-switching and time-varying growths, *Journal of Business and Economic Statistics*, **17**(4), 456–465.

Maheu, J.M. and McCurdy, T.H. (2000), Identifying bull and bear markets in stock returns, *Journal of Business and Economic Statistics*, **18**(1), 100–112.

Raymond, J.E. and Rich, R.W. (1997), Oil and the macroeconomy; a Markov state-switching approach, *Journal of Money, Credit, and Banking*, **29**(2), 193–213.

Scheicher, M., (1999), Nonlinear dynamics: evidence for a small stock exchange, *Empirical Economics*, **24**, 45–59.

So, M.K.P., Lam, K. and Li, W.K. (1998), A stochastic volatility model with Markov switching, *Journal of Business and Economic Statistics*, **16**(2), 244–253.

Town, R.J. (1992), Merger waves and the structure of merger and acquisition time series, *Journal of Applied Econometrics*, **7**(supplement), S83–100.

8 Present value models and tests for rationality and market efficiency

The present value model states that the present value of an asset is derived from its earning power, or the ability to generate future income. This crucially depends on the expectations about future income and the discount rate at which people or investors would sacrifice a portion of their current income for future consumption, after adjusting for uncertainty or risk involved in the process. Although the present value of an asset, or economic value, as against accounting value, is the best to reflect its true value, it involves expectations on future income, the discount rate and rationality of people. Therefore the present value model is difficult to apply properly in practice. To link the present value of an asset to its future income in the framework of cointegration analysis, as proposed by Campbell and Shiller (1987), has provided a useful tool for testing expectations and rationality in financial markets.

8.1. The basic present value model and its time series characteristics

The present value of an asset is its all future income discounted:

$$V_t = \sum_{\tau=1}^{\infty} \frac{1}{(1+r_t)\cdots(1+r_{t+\tau})} E_t I_{t+\tau} \tag{8.1}$$

where V_t is the present value of the asset, I_{t+1} is income derived from possessing this asset in period $(t, t+1]$, E_t is expectations operator, and r_t is the discount rate in period $(t, t+1]$.

When the discount rate is constant, i.e., $r_t = r$, equation (8.1) becomes:

$$V_t = \sum_{\tau=1}^{\infty} \frac{1}{(1+r)^\tau} E_t I_{t+\tau} \tag{8.1'}$$

Subtracting $V_t/(1+r)$ from both sides leads to:

$$V_t - \frac{V_t}{1+r} = \sum_{\tau=1}^{\infty} \frac{1}{(1+r)^\tau} E_t I_{t+\tau} - \sum_{\tau=1}^{\infty} \frac{1}{(1+r)^\tau} E_t I_{t+\tau-1} + \frac{I_t}{1+r}$$

Re-arrangement of the above yields:

$$V_t - \frac{I_t}{r} = \frac{1+r}{r} \sum_{\tau=1}^{\infty} \frac{1}{(1+r)^\tau} E_t \Delta I_{t+\tau} \tag{8.2}$$

Equation (8.2) states that if V_t and I_{t+1} are $I(1)$ series, then a linear combination of them is stationary too and the two series are cointegrated. Campbell and Shiller (1987) define $(V_t - I_t)/r$ as spread, S_t. Obviously, the spread links a stock variable, V_t, to a flow variable, I_t. It is not strange that a flow variable divided by the rate of flow (in this case r) is a stock variable; or a stock variable times the rate of flow is a flow variable. If income is constant over time, then total wealth or value of an asset is simply the current income flow divided by the rate at which income is generated, i.e., the spread is equal to zero. Otherwise, the spread is a function of the expected changes in future incomes discounted. A positive spread reflects an overall growth in future incomes, and a negative spread is associated with income declines.

Nevertheless, the seemingly stationarity of the right-hand side in equation (8.2) is problematic, or at least unrealistic. The growth or change in income as expressed in equation (8.2) is in an absolute term, $I_t - I_{t-1}$, instead of a relative term, $(I_t - I_{t-1})/I_{t-1}$. Let us adopt a version of the Gordon dividend growth model:

$$V_t = \sum_{\tau=1}^{\infty} \frac{E_t I_{t+\tau}}{(1+r)^\tau} = \sum_{\tau=1}^{\infty} \frac{(1+g)^\tau I_t (1+E_t u_{t+\tau})}{(1+r)^\tau} \tag{8.3}$$

subtracting $(1+g)V_t/(1+r)$ from both sides, we have:

$$V_t - \frac{(1+g)V_t}{1+r} = \sum_{\tau=1}^{\infty} \frac{(1+g)^\tau I_t (1+E_t u_{t+\tau})}{(1+r)^\tau} - \sum_{\tau=1}^{\infty} \frac{(1+g)^\tau I_t (1+E_t u_{t+\tau-1})}{(1+r)^\tau}$$
$$+ \frac{(1+g)I_t}{1+r}$$

and re-arrangement yields:

$$V_t - \frac{(1+g)I_t}{r-g} = \frac{1+r}{r-g} \sum_{\tau=1}^{\infty} \frac{(1+g)^\tau E_t \Delta u_{t+\tau}}{(1+r)^\tau} \tag{8.4}$$

where $\Delta u_{t+\tau} = \Delta \ln I_{t+\tau} - g$. Equation (8.4) reduces to the Campbell and Shiller (1987) formulation when $g = 0$. ΔI_t is in general non-stationary.

Define $V_t - [(1+g)I_t/(r-g)]$ as the full spread, Sf_t. Equation (8.4) says total wealth or value of an asset is simply the current income flow (notice $I_{t+1} = (1+g)I_t$ is the income in the current period) divided by, instead of the discount rate, the difference between the discount rate and the growth rate. Equation (8.4) is, in fact, the Gordon valuation model for constant growing perpetuities. The important

message here is that there exists a cointegration or long-run relationship between the value and the income, as revealed by equations (8.2) and (8.4). Moreover, if income obeys a constant growth process, the spread in the sense of Campbell and Shiller (1987) is not stationary, but the full spread as defined above is stationary. Therefore, caution has to be taken in explaining and interpreting the cointegration vector. If $(1+g)/(r-g)$ is mistaken as $1/r$, then $(r-g)/(1+g)$ might be mistaken as the discount rate r and the practice would *under* estimate the true discount rate if there is growth in income.[1] Later in this chapter, we will see how to impose restrictions and carry out empirical tests.

Equation (8.2) can also be written as:

$$S_t = V_t - \frac{1}{r}I_t = \frac{1+r}{r}E_t\Delta V_{t+1} \tag{8.5}$$

If there is a rational bubble b_t, satisfying:

$$b_t = \frac{1}{1+r}E_t b_{t+1}$$

i.e.:

$$b_{t+1} = (1+r)b_t + \zeta_{t+1}, \ \zeta_t \sim \text{iid}(0, \sigma_\zeta^2) \tag{8.6}$$

in equation (8.1′), it will appear on the right-hand side of equation (8.2), but will not appear on the right-hand side of equation (8.5). b_t has a root outside the unit circle and is explosive or non-stationary. Consequently, even if ΔI_t is stationary, the spread is non-stationary if there is a rational bubble in equation (8.1′), inducing non-stationarity in ΔV_t through equation (8.5). Therefore, testing for rationality is equivalent to testing for cointegration between the present value variable, V_t, and the income variable, I_t.

8.2. The VAR representation

Equations (8.2) and (8.5) also suggest a way to compute the variables in a VAR (I assume I had introduced such terminologies in earlier chapters). Let $\mathbf{z}_t = \begin{bmatrix} S_t & \cdots & S_{t-p+1} & \Delta I_t & \cdots & \Delta I_{t-p+1} \end{bmatrix}'$, the VAR can be written in the companion form:

$$\begin{bmatrix} \mathbf{S}_t \\ \Delta \mathbf{I}_t \end{bmatrix} = \begin{bmatrix} \mathbf{a}_{11}(L) & \mathbf{a}_{12}(L) \\ \mathbf{a}_{21}(L) & \mathbf{a}_{22}(L) \end{bmatrix} \begin{bmatrix} \mathbf{S}_{t-1} \\ \Delta \mathbf{I}_{t-1} \end{bmatrix} + \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} \tag{8.7}$$

where

$$\mathbf{S}_t = \begin{bmatrix} S_t & \cdots & S_{t-p+1} \end{bmatrix}', \qquad \Delta \mathbf{I}_t = \begin{bmatrix} \Delta I_t & \cdots & \Delta I_{t-p+1} \end{bmatrix}'$$

$$\mu_{1t} = \begin{bmatrix} v_{1t} & 0 & \cdots \end{bmatrix}', \qquad \mu_{2t} = \begin{bmatrix} v_{2t} & 0 & \cdots \end{bmatrix}'$$

$$\mathbf{a}_{11}(L) = \begin{bmatrix} a_{11,1} & \cdots & \cdots & a_{11,p} \\ 1 & \cdots & & \\ & & \cdots & \\ & & & 1 \end{bmatrix}, \qquad \mathbf{a}_{12}(L) = \begin{bmatrix} a_{12,1} & \cdots & \cdots & a_{12,p} \\ 0 & \cdots & & \\ & & \cdots & \\ & & & 0 \end{bmatrix}$$

$$\mathbf{a}_{21}(L) = \begin{bmatrix} a_{21,1} & \cdots & \cdots & a_{21,p} \\ 0 & \cdots & & \\ & & \cdots & \\ & & & 0 \end{bmatrix}, \qquad \mathbf{a}_{22}(L) = \begin{bmatrix} a_{22,1} & \cdots & \cdots & a_{22,p} \\ 1 & \cdots & & \\ & & \cdots & \\ & & & 1 \end{bmatrix}$$

Or, in a compact form:

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mu_t \tag{8.8}$$

The implication of this representation is that the spread, S_t, must linearly Granger cause ΔI_t, unless S_t is itself an *exact* linear combination of current and lagged ΔI_t. Therefore, S_t would have incremental predicting power for ΔI_t. Further, let $\mathbf{e}1'$ and $\mathbf{e}2'$ be $(1 \times 2p)$ row vectors with zero in all cells except unity in the first element for the former and in the $(p+1)$ element for the latter, respectively, i.e.:

$$S_t = \mathbf{e}1'\mathbf{z}_t \tag{8.9}$$

$$\Delta I_t = \mathbf{e}2'\mathbf{z}_t \tag{8.10}$$

Notice:

$$E\{\mathbf{z}_{t+\mathbf{k}}|H_t\} = \mathbf{A}^k\mathbf{z}_t \tag{8.11}$$

where H_t is the information set with all available information about S_t and ΔI_t at time t. Applying equations (8.9), (8.10) and (8.11) to (8.2) yields:

$$\mathbf{e}1'\mathbf{z}_t = \frac{1}{r}\sum_{i=1}^{\infty}\frac{1}{(1+r)^i}\mathbf{e}2'\mathbf{A}^i\mathbf{z}_t = \mathbf{e}2'\frac{1}{r(1+r)}\mathbf{A}\left[\mathbf{I} - \frac{1}{1+r}\mathbf{A}\right]^{-1}\mathbf{z}_t \tag{8.12}$$

Equation (8.12) imposes restrictions on the VAR parameters if rationality is to hold, i.e.:

$$\mathbf{e}1'\left[\mathbf{I} - \frac{1}{1+r}\mathbf{A}\right] = \mathbf{e}2'\frac{1}{r(1+r)}\mathbf{A} \tag{8.13}$$

Accordingly, the 'theoretical' spread can be introduced as:

$$S_t^* = \mathbf{e}2'\frac{1}{r(1+r)}\mathbf{A}\left[\mathbf{I} - \frac{1}{1+r}\mathbf{A}\right]^{-1}\mathbf{z}_t \tag{8.14}$$

It can be seen that the difference between the actual and 'theoretical' spreads is:

$$S_t - S_t^* = \sum_{i=1}^{\infty} \frac{1}{(1+r)^i} E(\xi_{t+i} | H_t) \tag{8.15}$$

where $\xi_t = V_t - \left[(1+r)V_{t-1} - I_t\right] = V_t - E_{t-1}V_t$ is the innovation in forecasting V_t. Testing the restrictions in equation (8.13) is equivalent to testing that the right-hand side of equation (8.15) is just white noise with a mean of zero. Also, using volatility test, the variance ratio var(S_t)/var(S_t^*) should not be significantly larger than unity if the present value model is to hold. In addition, volatility test can be carried out with the innovation ξ_t and the innovation in the expected present value:

$$\xi_t^* \equiv \frac{1}{r} \sum_{i=0}^{\infty} \frac{1}{(1+r)^i} \left[E(\Delta I_{t+i} | H_t) - E(\Delta I_{t+i} | H_{t-1}) \right]$$

$$= S_t^* - (1+r)S_{t-1}^* + \frac{1}{r}\Delta I_t \tag{8.16}$$

The variance ratio var(ξ_t)/var(ξ_t^*) can be viewed as the 'innovation variance ratio', and var(S_t)/var(S_t^*) as the 'level variance ratio'. Notice ξ_t can also be written in the similar expression of equation (8.16):

$$\xi_t \equiv V_t - \left[(1+r)V_{t-1} - I_t\right]$$

$$= V_t - \frac{1}{r}I_{t+1} + \frac{1}{r}I_{t+1} - (1+r)V_{t-1} + (1+r)I_t - \frac{1}{r}I_t$$

$$= S_t - (1+r)S_{t-1} + \frac{1}{r}\Delta I_{t+1} \tag{8.17}$$

The implications of the above equations can be summarised in the following. If the market is rational for an asset, then its value/price and income variables should be cointegrated and its spread should be stationary. Without a cointegration relation between the price and income, the spread is non-stationary and a 'rational bubble', which by definition is explosive, would exist in the market. If the market is efficient and the present value model holds, then the 'theoretical' spread should not systematically differ from the actual spread, and both variance ratios should not be significantly larger then unity. The prediction power of the spread for ΔI_t is conditional on agents' information set. If agents do not have information useful for predicting ΔI_t beyond the history of ΔI_t, then S_t is a linear combination of current and lagged ΔI_t without prediction ability. Prediction may or may not be improved simply because the price and income variables are cointegrated. Therefore, in this chapter, we use cointegration between the price and income as a criterion for rationality against the existence of bubble in the market. In addition, we use the VAR representation and the variance ratios derived from the VAR system to examine whether the present value model holds and how far the market is from efficient.

8.3. The present value model in logarithms with time-varying discount rates

The previous section has shown that a ratio relationship between the value and income variables is more appropriate than a 'spread' relationship between the two variables, in the context of a constant discount rate and growth in income. As most economic and financial variables grow exponentially, linear relationships are only appropriate for variables in their logarithm, not for variables in their original form. This is equivalent to say that variables in their original form have ratio relationships, instead of linear relationships. In a sense, a right modelling strategy reflects impeccably both the economic and financial characteristics and the data generating process and makes these two considerations fit into each other. In this section, we further generalise the present value model along this line and allow for a time-varying rate of return or discount rate in the model.

In this section, we deliberate value, income and their relationship in a context of stock market investment explicitly, i.e., value and income variables are characterised by observable share prices and dividends. Let us express the rate of total return in the logarithm form:

$$r_t = \ln\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) \tag{8.18}$$

Notice r_t is an approximation of the exact rate of total return. However, this expression is in common with general practice and leads, conventionally, to the linear relationship among all variables involved.

As already known, total return can be split into price appreciation and the dividend yield. The idea is also valid in the log-linear form. To see this, expanding equation (8.18) as:

$$
\begin{aligned}
r_t &= \ln\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right) = \ln\left[\frac{P_{t+1}}{P_t}\left(1 + \frac{D_{t+1}}{P_{t+1}}\right)\right] = \ln\left(\frac{P_{t+1}}{P_t}\right) + \ln\left(1 + \frac{D_{t+1}}{P_{t+1}}\right) \\
&\approx (\ln P_{t+1} - \ln P_t) + \frac{D_{t+1}}{P_{t+1}} \\
&= (p_{t+1} - p_t) + e^{(d_{t+1} - p_{t+1})}
\end{aligned}
\tag{8.19}
$$

where, $p_t = \ln P_t$, and $d_t = \ln D_t$. The first term on the right-hand side is price appreciation, and the last term on the right-hand side reflects the dividend yield (notice the exact dividend yield is D_{t+1}/P_t). As the last term on the right-hand side is not linear, further transformation and approximation are required. Finally, after a series of development, the rate of total return can be expressed as:

$$r_t \approx \kappa + (1 - \lambda)p_{t+1} - p_t + \lambda d_{t+1} \tag{8.20}$$

where $\lambda = e^{(d-p)'} = (D/P)'$ is a constant between the minimum and maximum dividend yields, and $\kappa = (d - p)' e^{(d-p)'} = \ln(D/P)' \times (D/P)'$ is also a constant.

With the rate of total return, price and dividend being linked in a log-linear relationship as in equation (8.20), it is now possible to express the present value model in a log-linear form too. Furthermore, no restriction on the rate of return r_t to be constant is required to derive the log-linear form present value model. Thus, the model could accommodate the time-varying rate of return or discount rate and is more general and closer to reality.

Solving equation (8.20) forward, we obtain:

$$p_t = \frac{\kappa}{\lambda} + \sum_{\tau=0}^{T} \left\{ (1-\lambda)^\tau \left(\lambda d_{t+1+\tau} - r_{t+\tau} \right) \right\} + (1-\lambda)^T p_{T+1} \tag{8.21}$$

when $T \to \infty$, the last term on the right-hand side $\to 0$, and equation (8.21)

$$p_t = \frac{\kappa}{\lambda} + \sum_{\tau=0}^{\infty} (1-\lambda)^\tau \left(\lambda d_{t+1+\tau} - r_{t+\tau} \right) \tag{8.22}$$

Equation (8.22) is the log-linear counterpart of equation (8.1), and is not advantageous compared with the latter. Both are able to deal with the time-varying discount rate, but equation (8.1) is exact whereas equation (8.22) is an approximation. However, the benefit would be seen when the value–income or price–dividend relationship is examined. Extracting d_t from both sides of equation (8.22) and re-arrangement yield:

$$p_t - d_t = -(d_t - p_t) = \frac{\kappa}{\lambda} + \sum_{\tau=0}^{\infty} (1-\lambda)^\tau \left(\Delta d_{t+1+\tau} - r_{t+\tau} \right) \tag{8.23}$$

It can be observed that if d_t is $I(1)$ and r_t is $I(0)$, the left-hand side of equation (8.23) is also $I(0)$, or stationary. That is, the price and dividend in their logarithm are cointegrated. Notice no conditions are placed on r_t to derive the cointegration relationship, in contrast with equation (8.2). This is obviously advantageous, compared with the 'spread' form specification.

Equations (8.22) and (8.23) are derived as *ex post*, but they also hold *ex ante*. Taking expectations operations on both sides of equations (8.22) and (8.23), we have:

$$p_t = \frac{\kappa}{\lambda} + E_t \left\{ \sum_{\tau=0}^{\infty} (1-\lambda)^\tau \left(\lambda d_{t+1+\tau} - r_{t+\tau} \right) \right\} \tag{8.22'}$$

and:

$$p_t - d_t = -(d_t - p_t) = \frac{\kappa}{\lambda} + E_t \left\{ \sum_{\tau=0}^{\infty} (1-\lambda)^\tau \left(\Delta d_{t+1+\tau} - r_{t+\tau} \right) \right\} \tag{8.23'}$$

Previously in section 8.1, we have shown that value (price) and income (dividend) would be cointegrated with a cointegration vector $[1, -1/r]$, if the *absolute* changes

in income are stationary or constant. If the income stream has a constant growth rate, instead of constant absolute increase, then they would be cointegrated with a cointegration vector $[1, -1/(r-g)]$. Recall the derivation of a cointegration relationship is dependent on $r_t \equiv r$, so the cointegration relationship is rather restrictive. With the log-linear form present value model, the cointegration vector is always $[1, -1]$. The proportional relation for the price and dividend is reflected by the constant and variables on the right-hand side of equation (8.23) or equation (8.23′), which are time-varying in general. The cointegration between price and dividend is not affected by whether the discount rate is assumed to be constant or not as in section 8.1. As we know prices, dividends and most financial variables grow exponentially, there should be a log-linear relationship among them. Consequently, models in the log-linear form are generally sound, financially and statistically.

8.4. The VAR representation for the present value model in the log-linear form

The VAR representation of the log-linear form is similar to that of the original form. Let $\mathbf{z}_t = \begin{bmatrix} s_t & \dots & s_{t-p+1} & r_t - \Delta d_{t+1} & \dots & r_{t-p} - \Delta d_{t-p+1} \end{bmatrix}'$, where $s_t = d_t - p_t$. s_t is, roughly, the log dividend yield (the exact log divided yield is $d_{t+1} - p_t$). Compared with section 8.2, the spread S_t is replaced by the log-dividend yield, and the absolute changes in dividends are replaced by the difference between the percentage changes in dividends and the discount rate (Recall, in sections 8.1 and 8.2, r_t is restricted to a constant and did not appear in the $\mathbf{z}_t$ vector). With the same $\mathbf{A}$ matrix as in section 8.2, the compact form is:

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mu_t \tag{8.8'}$$

The selecting vector $\mathbf{e}1$ picks up s_t from $\mathbf{z}_t$ and the following holds, conditional on H_t, the information in the VAR:

$$s_t = \mathbf{e}1'\mathbf{z}_t = \sum_{\tau=0}^{\infty} (1-\lambda)^{\tau} \, \mathbf{e}2'\mathbf{A}^{\tau+1}\mathbf{z}_t = \mathbf{e}2'\mathbf{A}\left[\mathbf{I} - (1-\lambda)\mathbf{A}\right]^{-1}\mathbf{z}_t \tag{8.24}$$

Therefore:

$$\mathbf{e}1' = \mathbf{e}2'\mathbf{A}\left[\mathbf{I} - (1-\lambda)\mathbf{A}\right]^{-1} \tag{8.25}$$

or:

$$\mathbf{e}1'\left[\mathbf{I} - (1-\lambda)\mathbf{A}\right] - \mathbf{e}2'\mathbf{A} = 0 \tag{8.26}$$

The log dividend yield satisfying the conditions in equation (8.25) or equation (8.26) is the theoretical log dividend yield, written as s_t^*. Notice again there are no restrictions imposed on r_t, so tests on the validity of the present value model

in the log-linear form are not subject to the assumption about the discount rate. That is, the present value model can be accepted or rejected no matter the discount rate is treated as time-varying or not. The variance ratio test on s_t^* and s_t can be carried out to examine whether the present value model holds.

Furthermore, dividend volatility and return volatility can also be tested, respectively. If the discount rate is constant over time, it could be excluded from the $\mathbf{z}_t$ vector, and the theoretical log dividend yield with a constant discount rate, $s_{d,t}^*$, is obtained. The hypothesis for a constant discount rate is $H_{r0}: s_{d,t}^* = s_t^*$. In a separate study, Campbell and Shiller (1989) reject constant discount rate in the US stock market, employing the Cowles/S&P data set (1871–1986) and the NYSE data set (1926–1986). In a similar way, if $\Delta d_t = g$, i.e., the dividend growth is constant, then Δd_t can be excluded from the $\mathbf{z}_t$ vector too, and the theoretical log dividend yield with the constant dividend growth, $s_{r,t}^*$, emerges. The hypothesis for dividend growth to be constant is $H_{d0}: s_{r,t}^* = s_t^*$, though it has little financial meaning. The variance ratio test can also be employed to test these two hypotheses.

8.5. Variance decomposition

As returns may be volatile, we are interested in the sources of volatility. Substituting equation (8.22′) into equation (8.20) yields an expression for innovation in the total rate of return:

$$
r_t - E_t\{r_t\} = E_{t+1}\left\{\sum_{\tau=0}^{\infty}(1-\lambda)^{\tau}\,\Delta d_{t+1+\tau}\right\} - E_t\left\{\sum_{\tau=0}^{\infty}(1-\lambda)^{\tau}\,\Delta d_{t+1+\tau}\right\}
$$

$$
- \left[E_{t+1}\left\{\sum_{\tau=1}^{\infty}(1-\lambda)^{\tau}\,r_{t+\tau}\right\} - E_t\left\{\sum_{\tau=1}^{\infty}(1-\lambda)^{\tau}\,r_{t+\tau}\right\}\right] \quad (8.27)
$$

Equation (8.27) can be written in compact notations, with the left-hand side term being v_t, the first term on the right-hand side $\eta_{d,t}$, and the second term on the right-hand side $\eta_{r,t}$:

$$
v_t = \eta_{d,t} - \eta_{r,t} \tag{8.28}
$$

where v_t is the innovation or shock in total returns, $\eta_{d,t}$ represents the innovation due to changes in expectations about future income or dividends, and $\eta_{r,t}$ represents the innovation due to changes in expectations about future discount rates or returns.

Again, we use VAR to express the above innovations. Vector $\mathbf{z}_t$ contains, first of all, the rate of total return or discount rate. Other variables included are relevant to forecast the rate of total return:

$$
\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \varepsilon_t \tag{8.29}
$$

with the selecting vector $\mathbf{e}1$ which picks out r_t from $\mathbf{z}_t$, we obtain:

$$
v_t = r_t - E_t\{r_t\} = \mathbf{e}1'\varepsilon_t \tag{8.30}
$$

Bring equations (8.29) and (8.30) into the second term on the right-hand side of equation (8.27) yields:

$$
\eta_{r,t} = E_{t+1} \left\{ \sum_{\tau=1}^{\infty} (1-\lambda)^{\tau} r_{t+\tau} \right\} - E_t \left\{ \sum_{\tau=1}^{\infty} (1-\lambda)^{\tau} r_{t+\tau} \right\}
$$

$$
= \mathbf{e}1' \sum_{\tau=1}^{\infty} (1-\lambda)^{\tau} \mathbf{A}^{\tau} \varepsilon_t = \mathbf{e}1'(1-\lambda)\mathbf{A} \left[\mathbf{I} - (1-\lambda)\mathbf{A} \right]^{-1} \varepsilon_t \qquad (8.31)
$$

While $\eta_{d,t}$ can be easily derived according to the relationship in equation (8.28) as follows:

$$
\eta_{d,t} = v_t + \eta_{r,t} = \mathbf{e}1' \left\{ \mathbf{I} + (1-\lambda)\mathbf{A} \left[\mathbf{I} - (1-\lambda)\mathbf{A} \right]^{-1} \right\} \varepsilon_t \qquad (8.32)
$$

The variance of innovation in the rate of total return is the sum of the variance of $\eta_{r,t}$, innovation due to changes in expectations about future discount rates or returns, $\eta_{d,t}$, innovation due to changes in expectations about future income or dividends, and their covariance, i.e.:

$$
\sigma_v^2 = \sigma_{\eta,d}^2 + \sigma_{\eta,r}^2 - 2\mathrm{cov}(\eta_{d,t}, \eta_{r,t}) \qquad (8.33)
$$

8.6. Examples and cases

The present value model discussed in this chapter has provided a powerful approach to modelling value–income or price–dividend relationships via exploiting their time series characteristics, namely, cointegration and restrictions on the VAR. In this section, several examples in financial markets and international economics and finance are presented to illustrate how the research is empirically carried out.

Example 8.1

This is a case in US stock market behaviour in Campbell and Shiller (1987). The price and dividend data were of annual frequency from 1971 to 1986 in a broad stock index mainly represented by Standard and Poor's with adjustments. The model used was in the original form, i.e., without logarithm operations. The main results are summarised in Tables 8.1 and 8.2.

   The unit root test, which uses one of the Perron–Phillips test statistics, confirms that the stock price and dividend are $I(1)$ variables. The spread is stationary, when it is calculated with a discount rate of 3.2 per cent estimated with cointegration, but the spread is non-stationary when a discount rate of 8.2 per cent from the sample mean is applied. Based on these results, Campbell and Shiller suggest that a 'rational bubble' is not present but the

Table 8.1 Tests of stationarity, cointegration and rationality

	With trends	*Without trends*
I_t	−2.88	−1.28
V_t	−2.19	−1.53
ΔI_t	−8.40***	−8.44***
ΔV_t	−9.91***	−9.96***
$S_t = (V_t - 1/0.032 I_t)$	−4.35***	−4.31***
$r = 3.2\%$		
$S_t (= V_t - 1/0.082 I_t)$	−2.68	−2.15
$r = 8.2\%$		

***reject the null of a unit root at the 1 per cent level.
V_t represents the stock price variable and I_t represents the dividend variable.

Table 8.2 Tests of the present value model

	VAR restrictions	$var(S_t)/var(S_t^*)$	$var(\xi_t)/var(\xi_t^*)$
$r = 3.2\%$	5.75 (0.218)	4.786 (5.380)	1.414 (0.441)
$r = 8.2\%$	15.72 (0.0047)	67.22 (86.04)	11.27 (4.49)

p-value in parentheses for testing VAR restrictions which obeys the χ^2 distribution. Standard errors in parentheses for variance ratio tests.

evidence for cointegration between the stock price and dividend is weak as the stationarity of the spread is rejected if a 'more reasonable' discount rate is used. However, as has been pointed in section 8.1, the stock price and dividend, if cointegrated, will not always be cointegrated at $[1, -1/r]$. With growth in dividends, they are more likely to be cointegrated at $[1, -1/(r-g)]$, and an estimate of 3.2 per cent for $(r-g)$ may not be too low. Therefore, the estimate should be interpreted as $(r-g)$ instead of r.

Although the US stock market is not subject to a 'rational bubble' and the stock market behaviour is rational, the present value model may not hold. This is examined by testing variance ratios of the unrestricted and theoretical specifications, and imposing restrictions on the VAR and testing for their validity. Selected testing statistics in Table 8.2 suggest the present value model is rejected for the US stock market. The variance ratio test statistics are greater than unity, though only the innovation variance ratio is statistically significant. Tests for VAR restrictions in equation (8.13) accepts the model with the 3.2 per cent discount rate and rejects it with the 8.2 per cent discount rate. As mentioned in the above, the US stock price and dividend during this period are more likely to be cointegrated at $[1, -1/(r-g)]$,

Continued

so when the discount rate from the sample mean is applied, the right-hand side of equation (8.2) may not be stationary, or in fact, it is the right-hand side of equation (8.3). So, the mixed results tilt to imply the validity of the VAR model. As expected, with the 8.2 per cent discount rate, the variance ratios are much greater than unity, but with very large standard errors.

Example 8.2

This is an example of the present value model's applications in the real estate market. The data used are capital value and rental indices from Jones Lang Wootten (JLW). The JLW index is one of the major UK real estate indices. The data sets are of quarterly frequency from the second quarter in 1977 to the first quarter in 1997, at the aggregate level as well as the disaggregate level for office, industrial and retail sectors. After confirming both capital value and rent variables are $I(1)$ series, cointegration between the capital value and the rent, or stationarity of the spread, is examined. The study uses the Johansen procedure for testing the cointegration relationship. Although there are only two variables, it is beneficial to use the Johansen procedure in a dynamic setting. The cointegration test is carried out with the variables in their original form and in logarithm, the latter is able to deal with a time-varying rate of return or discount rate. However, the two sets of results in Tables 8.3 and 8.4 are virtually the same, implying the model in the original form is acceptable in this case. The results suggest that there are no bubbles in the office, retail and aggregate property markets; but the existence of bubbles in the industrial property market cannot be ruled out. The industrial property is probably the most illiquid and indivisible among all types of property, and as a consequence, its price/capital value fails

Table 8.3 Check for stationarity of S_t-cointegration of V_t and I_t

	λ_{max}	λ_{trace}
Office	25.66***	28.63***
Industrial	12.29	18.54
Retail	16.83	25.20*
All	34.15***	38.23***

Model with unrestricted constant and restricted trend. Lag lengths are selected with a compromise of the Akaike, Schwarz and Hannan-Quinn criteria.
*reject zero cointegration vector (accept one cointegration vector) at the 10 per cent level;
***reject zero cointegration vector (accept one cointegration vector) at the 1 per cent level.
Critical values from Osterwald-Lenum (1992). Critical values for one cointegration vector are: for λ_{max}: 16.85 (90 per cent), 28.96 (95 per cent) and 23.65 (99 per cent); for λ_{trace}: 22.76 (90 per cent), 25.32 (95 per cent) and 30.45 (99 per cent); V_t represents capital value and I_t represents rent.

Table 8.4 Check for stationarity of s_t-cointegration between the logarithm of V_t (v_t) and the logarithm of I_t (i_t)

	λ_{max}	λ_{trace}
Office	24.16***	26.98**
Industrial	14.67	17.66
Retail	14.81	23.88*
All	28.67***	33.11***

See notes in Table 8.3.
*reject zero cointegration vector (accept one cointegration vector) at the 10 per cent level;
**reject zero cointegration vector (accept one cointegration vector) at the 5 per cent level;
***reject zero cointegration vector (accept one cointegration vector) at the 1 per cent level.

to reflect its future income in transactions. Though this phenomenon is generally ruled as the existence of bubbles, it should not simply be made equal to speculation. A 'thin' market for industrial property transactions may reasonably explain a large part of this particular statistical result for the industrial property.

The validity of the present value model is examined and the testing statistics are reported in Tables 8.5 and 8.6. The validity of the VAR model is rejected in all types of properties except the office property. In general, the spread causes the change in the rent, implying that the spread can help predict future rent; but changes in rent do not cause the spread in the aggregate and industrial properties, and they cause the spread in the office and retail properties at a lower significant level. The rejection of the VAR model is also reflected in Table 8.6 for variance ratio tests. The ratio of the variance of the spread to that of the 'theoretical' spread, i.e., the 'levels variance

Table 8.5 Tests with the VAR model

	S_t causes ΔI_t	ΔI_t causes S_t	Restrictions on VAR	Q(18)	
				ΔI_t	S_t
Office	17.6613***	3.3616**	1.1121	15.2080	14.2482
Industrial	9.4856***	1.5721	2.5610***	17.9259	17.7638
Retail	11.3278***	3.2626**	3.5616***	13.8085	10.6911
All	23.3608***	2.0033	1.9635**	16.3478	11.8133

**significant at the 5 per cent level; ***significant at the 1 per cent level.
Test statistics are F-test for causality test and restrictions on the VAR model, with respective degrees of freedom.
Q(18) is the Ljung–Box statistic for serial correlation up to 18 lags, in the rent equation (ΔI_t) and the spread equation (S_t) respectively.

Continued

Table 8.6 Variance ratios

	var(S_t)/var(S_t^*)	var(ξ_t)/var(ξ_t^*)
Office	$144.82/66.29 = 2.1846$***	$73.96/39.95 = 1.8513$***
Industrial	$273.90/24.64 = 11.1161$***	$59.36/27.74 = 2.1399$***
Retail	$512.83/113.00 = 4.538$***	$207.76/53.07 = 3.9148$***
All	$184.15/85.19 = 2.1616$***	$67.34/33.23 = 2.0265$***

*significantly different at the 10 per cent level; ***significantly different at the 1 per cent level.

ratio', is statistically significant in all types of properties. The 'innovation variance ratio' is also significant in all the cases, but the value is usually smaller. Observing these statistical numbers in detail, it is found that the office property is the least inefficient with the smallest 'innovation variance ratio'; and the industrial and retail properties are the most inefficient. This phenomenon is also reflected in the cointegration test, where capital value and rent are not cointegrated for the industrial property, and capital value and rent are cointegrated at a less significant level of 10 per cent for the retail property judged by the value of λ_{trace}.

Example 8.3

The present value model with cointegration can also be applied to international economics and finance. An example is in MacDonald and Taylor (1993) on the monetary model for exchange rate determination. We need some transformation before getting the present value representation for the monetary model. The flexible price monetary model is based on the following three equations:

$$m_t' - p_t' = \gamma y_t' - \lambda i_t' \tag{8.34}$$

$$s_t = p_t' \tag{8.35}$$

$$i_t' = E_t\left(\Delta s_{t+1}\right) = E_t\left(s_{t+1}\right) - s_t \tag{8.36}$$

where m_t' is money supply, p_t' is price level, y_t' is income, i_t' is interest rate, s_t is the exchange rate, γ is the income elasticity, and λ is the interest rate (semi) elasticity. All variables are in the logarithm; and except the exchange rate, all variables are the difference between the domestic variable and the foreign variable, e.g., $p_t' = p_t - p_t^*$. Equation (8.34) states a relative money market equilibrium requirement, equation (8.35) is the PPP (Purchasing Power Parity) condition, and equation (8.36) is the UIP (Uncovered

Interest rate Parity). Let $x_t = m'_t - \gamma y'_t$, then the exchange rate can be expressed as:

$$s_t = p'_t = x_t + \lambda i'_t = x_t + \lambda \left(E_t(s_{t+1}) - s_t \right) \tag{8.37}$$

or:

$$s_t = \frac{x_t}{1+\lambda} + \frac{\lambda}{1+\lambda} E_t(s_{t+1}) \tag{8.38}$$

Equation (8.38) can be solved forward and lead to:

$$s_t = \sum_{\tau=0}^{\infty} \frac{\lambda^\tau}{(1+\lambda)^{\tau+1}} E_t(x_{t+\tau+1}) \tag{8.39}$$

as:

$$\left(\frac{\lambda}{1+\lambda} \right)^T E_t(s_{t+1}) \to 0 \quad \text{when } T \to \infty$$

Equation has the same structure as the present value model, and it is easy to work out the following 'spread':

$$s_t - x_t = \sum_{\tau=0}^{\infty} \frac{\lambda^\tau}{(1+\lambda)^\tau} E_t(\Delta x_{t+\tau+1}) \tag{8.40}$$

Applying the same logic in section 8.1, the implication of equation (8.40) is that the exchange rate and x_t should be cointegrated and the 'spread' should be stationary, if the monetary model is to hold and a rational bubble does not exist in the foreign exchange market.

MacDonald and Taylor use the Johansen procedure (Johansen 1988 and Johansen and Juselius 1990) for the cointegration test, as $s_t - x_t$ involve five variables. Using the exchange rate data for the deutsche mark *vis-à-vis* the US dollar, they rule out rational bubbles in mark–dollar exchange market. However, with four estimates of λ, they firmly reject the VAR model with the imposed restrictions. These are summarised in Table 8.7.

Table 8.7 Tests of the VAR restrictions in the monetary model

	VAR restrictions	var(S_t)/var(S^*_t)
$\lambda = 0.050$	0.29 e + 07 (0.000)	0.11 e + 03 (0.000)
$\lambda = 0.030$	0.81 e + 07 (0.000)	0.30 e + 03 (0.000)
$\lambda = 0.015$	0.33 e + 08 (0.000)	0.12 e + 04 (0.000)
$\lambda = 0.001$	0.73 e + 10 (0.000)	0.29 e + 06 (0.000)

p-value in parentheses. Testing statistic for the VAR restrictions obeys χ^2 distribution. The variance ratio test employs the *F*-statistic.

Example 8.4

Previous cases have paid attention to the rationality in the market and the validity of the VAR representation of the present value model. The following example uses the present value model in the logarithm form to decompose variance or risk in returns. Liu and Mei (1994) apply the approach to the US equity Real Estate Investment Trusts (REITs) data. The data set is monthly and runs from January 1971 to December 1989. In addition to REITs, they included returns on value-weighted stock portfolio and on small stock portfolio in the VAR as other forecasting variables. The main results are summarised in Table 8.8.

The general message is that the variance of shocks in returns can be decomposed via the present value model in logarithm and the relative impacts of shocks or news in income (cash-flow risk as they called) and shocks or news in discount rates (discount-rate risk) can be assessed. Specifically, this study suggests that cash-flow risk accounts for a much larger proportion (79.8 per cent) in the total risk, compared with value-weighted stocks (38.1 per cent) and small stocks (29.7 per cent). As the correlation of the cash-flow shock and the discount-rate shock is positive, the total variance tends to decline when the two shocks are of the opposite signs (since the contribution of the covariance is $-2\text{cov}(\eta_{d,t}, \eta_{r,t})$). This is again different from value weighted stock portfolio which has a negative correlation between the two shocks, but is similar to small stocks. This study follows the paper 'A variance decomposition for stock returns' by Campbell (1991) which proposes and applies the approach to the US stock market.

Table 8.8 Variance decomposition for returns in REITs

	σ_ε^2	*Proportion of*			
		$\sigma_{\eta,d}^2$	$\sigma_{\eta,r}^2$	$-2\text{cov}(\eta_{d,t}, \eta_{r,t})$	$corr(\eta_{d,t}, \eta_{r,t})$
REITs	22.66	0.798	0.467	−0.265	0.217
		(0.40)	(0.40)	(0.66)	(0.41)
VWStk	21.97	0.381	0.333	0.386	−0.401
		(0.21)	(0.20)	(0.19)	(0.38)
SmStk	41.41	0.297	0.947	−0.244	0.230
		(0.13)	(0.52)	(0.61)	(0.47)

VWStk stands for return on value-weighted stock portfolio; SmStk stands for return on small stock portfolio. Standard errors in brackets. The VAR model was estimated with two lags and three lags respectively, only the former is reported here.

8.7. Empirical literature

The present value model links the (present) value of an asset to the future income or cash flows generated from possessing that asset in a fundamental way. Study of the validity of the present value model with cointegration analysis is a powerful method in empirical finance research. The analysis can be extended to investigate issues in bond markets, foreign exchange markets and other securities markets as well, where the relationships between the variables do not appear to be straightforward.

One of the important financial variables is the interest rate, which is central to the valuation of many other financial securities. As such, the term structure of interest rates, i.e., the relationships between long term and short term interest rates, or generally speaking, between interest rates of various maturities, have been a focus of study in a volatile financial investment environment. Applying the present value model, Veenstra (1999) investigates the relationship between spot and period freight rates for the ocean dry bulk shipping market, where the period rate is formulated as expectations of future spot rates. Formal tests on the VAR model reject the restriction imposed by the present value model. But the author argues that there is considerable evidence that the present value model is valid in ocean dry bulk shipping market after having considered alternative and informal test results. Nautz and Wolters (1999) test the expectations theory of the term structure, focusing on the question of how monetary policy actions indicated by changes in the very short rate affect long term interest rates. They claim that the expectations hypothesis implies that very long rates should only react to unanticipated changes of the very short rate, which only requires rational expectations but not stationary risk premia. That is, they challenge that there should be a cointegration relationship between very short rates and very long rates, and provide their explanation for the determinants of the term structure of interest rates.

There are a number of studies on foreign exchange rate determination similar to Example 8.3 in Section 8.6. Smith (1995) applies the present value model to formulate nominal exchange rates as discounted expected future fundamentals. The author rejects the validity of the present value relationship based on the findings that the discount rate obtained is statistically significantly negative. Nagayasu (1998) analyses Japanese long-run exchange rates using several exchange rate models, including the present value model. The author finds that the long run specification is sensitive to the specification of the model. A relevant area of study is how the current account is determined and influenced by the forcing variables. In this regard, Otto (1992) examines the post war data for the US and Canada, applying the present value relationship based upon the permanent income hypothesis of private consumption behaviour under rational expectations. The study strongly rejects the stringent restrictions imposed on the present value model with the US and Canadian data.

Research on the stock market remains one of the most active areas in which the present value model is empirically investigated, as the relationships between the

price and dividends are explicitly defined and stock market investment accounts for the largest amount of all types of investment in the world. Investigating stock prices on the Shanghai Stock Exchange, Chow *et al.* (1999) adopt the log-linear version of the present value model. Surprisingly, they find the model explains well the prices of 47 traded stocks as observed at the beginning of 1996, 1997 and 1998. There is some doubt cast on the use of such a data sample. Chow and Liu (1999) claim that stock prices can move in more volatile fashion than could be warranted by future dividend movements, when there is memory in the duration of dividend swings, if a constant discount rate is used in the present value model. The memory in the duration of a dividend swing will generate a spurious bias in the stock price and induce excess volatility in the stock price as if rational bubbles exist. More studies can be found in the papers by, for example, Crowder and Wohar (1998) and Lee (1995).

Due to the unique characteristics of low liquidity and high transaction costs, the behaviour of farmland and housing prices has been subject to extensive studies with regard to rationality and the existence of bubbles in the market. Bearing this in mind, Lence and Miller (1999) investigate whether the farmland 'constant-discount-rate present-value-model (CDR-PVM) puzzle' is due to transaction costs. They first discuss the theoretical implications of transaction costs for the CDR-PVM of farmland, then test the model with Iowa farmland prices and rents. Their empirical results regarding the validity of the CDR-PVM in the presence of typical transaction costs are not conclusive. Meese and Wallace (1994) examine the efficiency of residential housing markets by inspecting price, rent, and cost of capital indices generated from a transactions level data base for Alameda and San Francisco Counties in Northern California. They reject both constant and non-constant discount rate versions of the price present value model in the short term. Nevertheless, long-run results are consistent with the present value relation when they adjust the discount factor for changes in both tax rates and borrowing costs. Their explanation for the short term rejection and long run consistency is the high transaction costs in the housing market. Clayton (1996), Lloyd (1994) and Pindyck (1993) are also in this category of empirical research.

Note

1 Estimated with the cointegration regression, Campbell and Shiller (1987) reported a 3.2 per cent discount rate for the US broad stock market index from 1871 to 1986, which was substantially lower than the estimated mean rate of return of 8.2 per cent during this period. The difference in the two estimates, in fact, implies a 4.8 per cent growth in dividends.

Questions and problems

1 Why and how could the underlying economic processes and characteristics be better represented and reflected in an appropriate modelling strategy and framework? How could the consideration on statistics and the consideration on economics and finance be fitted into each other?

2   What are the advantages of linking value and income with the present value model in the original form? What are the shortcomings associated with this kind of modelling?

3   What are the advantages of linking value and income with the present value model in the logarithm form? Is modelling with the logarithm form an overall improvement over that with the original form, and why? Is it perfect?

4   It is often claimed that cointegration of two or more financial time series means market inefficiency. But in this chapter, cointegration between the price and dividend is a prerequisite for market efficiency, though it does not guarantee market efficiency. Explain.

5   Collect data from Datastream to test for cointegration between the price and dividend, using UK market indices:

   (a)   with the original data,
   (b)   data in logarithm.

6   Collect two companies' data from Datastream to test for cointegration between the price and dividend. One of the companies is a fast growing firm, and the other is rather stable. Again data are in the following forms:

   (a)   the original data,
   (b)   data in logarithm.

   Discuss the two sets of results you have obtained. Do they differ? Explain.

References

Campbell, J.Y. (1991), A variance decomposition for stock returns, *The Economic Journal*, **101**, 157–179.

Campbell, J.Y. and Shiller, R.J. (1987), Cointegration and tests of present value models, *Journal of Political Economy*, **95**, 1062–1088.

Campbell, J.Y. and Shiller, R.J. (1989), The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies*, **1**, 195–228.

Chow, G.C., Fan, Z.Z. and Hu, J.Y. (1999), Shanghai stock prices as determined by the present-value model, *Journal of Comparative Economics*, **27**(3), 553–561.

Chow, Y.F. and Liu, M. (1999), Long swings with memory and stock market fluctuations, *Journal of Financial and Quantitative Analysis*, **34**(3), 341–367.

Clayton, J. (1996), Rational expectations, market fundamentals and housing price volatility, *Real Estate Economics*, **24**(4), 441–470.

Crowder, W.J. and Wohar, M.E. (1998), Stock price effects of permanent and transitory shocks, *Economic Inquiry*, **36**(4), 540–552.

Johansen, S. (1988), Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control*, **12**(2/3), 231–254.

Johansen, S. and Juselius, K. (1990), Maximum likelihood estimation and inference on cointegration – with applications to the demand for money, *Oxford Bulletin of Economics and Statistics*, **52**(2), 169–210.

Lee, B.S. (1995), Fundamentals and bubbles in asset prices: evidence from U.S. and Japanese asset prices, *Financial Engineering and the Japanese Markets*, **2**(2), 89–122.

Lence, S.H. and Miller, D.J. (1999), Transaction costs and the present value model of farmland: Iowa, 1900–1994, *American Journal of Agricultural Economics*, **81**(2), 257–272.

Liu, C.H. and Mei, J. (1994), An analysis of real-estate risk using the present value model, *Journal of Real Estate Finance and Economics*, **8**, 5–20.

Lloyd, T. (1994), Testing a present value model of agricultural land values, *Oxford Bulletin of Economics and Statistics*, **56**(2), 209–223.

MacDonald, R. and Taylor, M. (1993), The monetary approach to the exchange rate – rational expectations, long-run equilibrium, and forecasting, *IMF Staff Papers*, **40**, 89–107.

Meese, R. and Wallace, N.E. (1994), Testing the present value relation for housing prices: should I leave my house in San Francisco?, *Journal of Urban Economics*, **35**(3), 245–266.

Nagayasu, J. (1998), Japanese effective exchange rates and determinants: prices, real interest rates, and actual and optimal current accounts, *International Monetary Fund Working Paper*: WP/98/86.

Nautz, D. and Wolters, J. (1999), The response of long-term interest rates to news about monetary policy actions: empirical evidence for the U.S. and Germany, *Weltwirtschaftliches Archiv*, **135**(3), 397–412.

Osterwald-Lenum, M. (1992), A note with quantiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics, *Oxford Bulletin of Economics and Statistics*, **54**, 461–472.

Otto, G. (1992), Testing a present-value model of the current account: evidence from the U.S. and Canadian time series, *Journal of International Money and Finance*, **11**(5), 414–430.

Pindyck, R. (1993), The present value model of rational commodity pricing, *Economic Journal*, **103**, 511–530.

Smith, G.W. (1995), Exchange rate discounting, *Journal of International Money and Finance*, **14**, 659–666.

Veenstra, A.W. (1999), The term structure of ocean freight rates, *Maritime Policy and Management*, **26**(3), 279–293.

9 State space models and the Kalman filter

A dynamic system can be described by changes in the state of its components. The variables of concern, which are observable, are represented as dynamic functions of these components, which are unobservable. The unobserved components, also called state variables, transit from one state to another or evolve according to certain rules which are not easy or straightforward to be applied to the observed variables themselves. This kind of dynamic modelling of systems is called the state space method. It explains the behaviour of externally observed variables by examining the internal, dynamic and systematic properties of the unobserved components. Therefore, this modelling strategy, if applied properly, may reveal the nature and cause of dynamic movement of variables in an effective and fundamental way.

State space models can be estimated using the Kalman filter, named after Kalman (1960, 1963), which was originally for and is still widely used in automatic control and communications. Initial application results are in Kalman and Bucy (1961) and subsequent developments are summarised by Kalman (1978). Clark (1987) was among the first to apply the state space model, using the Kalman filter, to economic analysis. Harvey (1989) and Hamilton (1994) contain a substantial element of this modelling method.

9.1. State space expression

The state space representation of a dynamic system can be formulated as:

$$y_t = H\xi_t + Ax_t + \mu_t \tag{9.1}$$

$$\xi_{t+1} = F\xi_t + Bx_{t+1} + v_{t+1} \tag{9.2}$$

where y_t is an $n \times 1$ vector of observed variables, ξ_t is a $r \times 1$ vector of state variables, x_t is a $k \times 1$ vector of exogenous variables, and H, A and F are coefficient matrices of dimension $n \times r$, $n \times k$ and $r \times r$ respectively. μ_t and v_t are vectors of residuals of dimension $n \times 1$ and $r \times 1$, with the following covariance matrices:

$$\text{Cov}(\mu_t\mu_t) = R, \quad \text{Cov}(v_t v_t) = Q, \quad \text{Cov}(\mu_t v_t) = 0 \tag{9.3}$$

Equation (9.1) is the observation equation or measurement equation; and equation (9.2) is the state equation or transition equation. They can be estimated by the Kalman filter algorithm to be illustrated in the next section.

9.2. Kalman filter algorithms

The Kalman filter can be better demonstrated in three steps, though at least the first two steps can be easily combined. The three steps are prediction, updating, and smoothing.

9.2.1. Predicting

This step is to predict, based on information available at $t-1$, the state vector $\xi_{t|t-1}$, its covariance matrix $P_{t|t-1}$ and derive an estimate of y_t accordingly:

$$\xi_{t,t-1} = F\xi_{t-1|t-1} \tag{9.4}$$

$$P_{t,t-1} = FP_{t-1|t-1}F^{'} + Q \tag{9.5}$$

$$y_{t|t-1} = H\xi_{t|t-1} + Ax_{t|t-1} \tag{9.6}$$

9.2.2. Updating

At this stage, the inference about ξ_t is updated using the observed value of y_t:

$$\psi_t = HP_{t|t-1}H^{'} + R \tag{9.7}$$

$$K_t = P_{t|t-1}H^{'}(\psi_t)^{-1} \tag{9.8}$$

$$\varepsilon_t = y_t - y_{t|t-1} \tag{9.9}$$

$$\xi_{t,t} = \xi_{t|t-1} + K_t\varepsilon_t \tag{9.10}$$

$$P_{t,t} = (I - K_tH)P_{t|t-1} \tag{9.11}$$

where K_t is the Kalman filter gain, ψ_t can be regarded as the system wide variance/covariance matrix, and ε_t the system wide vector of residuals. Then estimation of the Kalman filter is straightforward. The conditional density function is:

$$f(y_t|I_{t-1}) = (2\pi)^{n/2}|\psi_t|^{1/2}\exp\left(-\frac{\varepsilon_t^{'}\psi_t\varepsilon_t}{2}\right) \tag{9.12}$$

where I_{t-1} is the information set at time $t-1$. The Kalman filter can be estimated by maximising the log likelihood of the density function

(ignoring the constant part):

$$\text{Max}: \sum_{t=1}^{T} \log f(y_t|I_{t-1}) = -\frac{nT}{2} \log(2\pi) + \text{Max}: -\frac{1}{2} \sum_{t=1}^{T} \left\{ \log(\psi_t) + \left(\varepsilon_t' \psi_t \varepsilon_t \right) \right\}$$
(9.13)

Estimated parameters and state variables can be obtained accordingly.

At the prediction stage, inference is made based on the information contained in state variables only. This inference, however, has to be revised, based on the realisation of, and interaction with, observable variables. This is done at the updating stage. State variables evolve in their own way and the filter is like a black box at the prediction stage. But the purposes of introducing state variables are estimation, presentation, and revelation of the governing stochastic process of y_t in an alternative, if not a better way. These can only be achieved by comparing the actual value of y_t and that predicted by state variables. Corresponding error correction is made to update state variables so they closely track the dynamic system. The linkage between state variables and observed variables is maintained this way.

9.2.3. Smoothing

The state variables estimated during the above two stages use all past information and the current realisation of y_t, not the whole sample information which includes future information that has not arrived at the time. For real time control and similar applications, these are all required and can be expected. For some other applications, however, it may be of interest to know the estimate of a state variable at any given time t, based on the whole information set up to the last observation at time T. This procedure is smoothing which updates state variables backwards instead of forwards from $T-1$:

$$\xi_{t,T} = \xi_{t|t} + V_t \left(\xi_{t+1|T} - \xi_{t+1|t} \right)$$
(9.14)

$$P_{t,T} = P_{t|t} + V_t \left(P_{t+1|T} - P_{t+1|t} \right) V_t'$$
(9.15)

where

$$V_t = P_{t|t} F' P_{t+1|t}^{-1}$$
(9.16)

$\xi_{t|T}$ is a inference of ξ_t based on the whole sample and $P_{t|T}$ is its covariance matrix. As the inference of the state variable vector and its covariance matrix at T, $\xi_{T|T}$ and $P_{T|T}$, is known from equations (9.10) and (9.11), all of $\xi_{t|T}$ and $P_{t|T}$ can be recursively obtained through equations (9.14)–(9.16).

9.3. Time-varying coefficient models

Previously, we use state variables as unobserved components of y_t, the observable economic or financial variables, in the analysis of dynamic systems. We can also

use state variables for other purposes to better describe a system or relax some untested restrictions in the formulation of the system. One of the most common restrictions is that coefficients of a model are constant. State space models can easily set up a dynamic model that let the coefficients time-vary. If we modify equations (9.1) and (9.2) as follows:

$$y_t = H(z_t)\xi_t + Az_t + \mu_t \tag{9.17}$$

$$\xi_{t+1} = F(z_t)\xi_t + v_{t+1} \tag{9.18}$$

That is, the matrices H and F, which are constant in equations (9.1) and (9.2), become functions of z_t, which includes lagged y_t and exogenous variables x_t. This treatment allows state variables ξ_t to be time-varying coefficients. Equation (9.17) is a usual regressional model except that its coefficients are time-varying. Equation (9.18) is the unobserved process governing the evolution of the coefficients.

A simplest time-varying coefficient model is to let ξ_t follow a random walk:

$$\xi_{t+1} = \xi_t + v_{t+1} \tag{9.19}$$

Other specifications include autoregressive processes so that the coefficients are mean-reverting. In all these case, $F(z_t)$ is just a constant.

9.4. State space models of commonly used time series processes

9.4.1. AR(p) process

$$y_t = c + v_t$$

$$v_t = \rho_1 v_{t-1} + \cdots + \rho_p v_{t-p} + \varepsilon_t \tag{9.20}$$

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

There are a few expressions, one of them is as follows. The observation equation is:

$$y_t = c + \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_t \\ v_{t-1} \\ \vdots \\ v_{t-p} \end{bmatrix} \tag{9.21}$$

The state equation is:

$$
\begin{bmatrix} \upsilon_{t+1} \\ \upsilon_t \\ \vdots \\ \upsilon_{t-p+1} \end{bmatrix} = \begin{bmatrix} \rho_1 & \rho_2 & \cdot & \cdot & \rho_p \\ 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ \cdot & & \cdot & & \\ 0 & \cdot & \cdot & 1 & 0 \end{bmatrix} \begin{bmatrix} \upsilon_t \\ \upsilon_{t-1} \\ \vdots \\ \upsilon_{t-p} \end{bmatrix} \tag{9.22}
$$

Therefore, the construction elements of the model are:

$$
y_t = y_t, \qquad \xi_t' = \begin{bmatrix} \upsilon_t & \upsilon_{t-1} & \cdots & \upsilon_{t-p} \end{bmatrix}, \qquad x_t = c
$$

$$
H = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}, \quad F = \begin{bmatrix} \rho_1 & \rho_2 & \cdot & \cdot & \rho_p \\ 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ \cdot & & \cdot & & \\ 0 & \cdot & \cdot & 1 & 0 \end{bmatrix}, \quad A = 1, \quad B = 0
$$

$$
\mu_t = 0, \qquad v_t' = \begin{bmatrix} \varepsilon_t & 0 & \cdots & 0 \end{bmatrix}
$$

$$
Q = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & 0 \\ \cdot & & \cdot & & \cdot \\ 0 & \cdot & \cdot & \cdot & 0 \end{bmatrix}, \qquad R = 0
$$

9.4.2. ARMA(p, q) process

$$
y_t = c + \rho_1 y_{t-1} + \cdots + \rho_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_{-q} \varepsilon_{t-q}
$$
$$
\varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{9.23}
$$

The observation equation is:

$$
y_t = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_q \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-q} \end{bmatrix} + \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_p \end{bmatrix} \begin{bmatrix} c \\ y_{t-1} \\ \vdots \\ y_{t-p} \end{bmatrix} \tag{9.24}
$$

The state equation is:

$$
\begin{bmatrix} \varepsilon_{t+1} \\ \varepsilon_t \\ \vdots \\ \varepsilon_{t-q+1} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ \cdot & & \cdot & & \\ 0 & \cdot & \cdot & 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \\ \vdots \\ \varepsilon_{t-q} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{9.25}
$$

The construction elements of the model are:

$$y_t = y_t, \qquad \xi_t' = \begin{bmatrix} \varepsilon_t & \varepsilon_{t-1} & \cdots & \varepsilon_{t-q} \end{bmatrix}, \qquad x_t' = \begin{bmatrix} c & y_{t-1} & \cdots & y_{t-p} \end{bmatrix}$$

$$H = \begin{bmatrix} 1 & \theta_1 & \cdots & \theta_q \end{bmatrix}, \qquad F = \begin{bmatrix} 0 & 0 & \cdot & \cdot & 0 \\ 1 & 0 & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ \cdot & \cdot & & & \\ 0 & \cdot & \cdot & 1 & 0 \end{bmatrix},$$

$$A = \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_p \end{bmatrix}, \qquad B = 0$$

$$\mu_t = 0, \qquad v_t' = \begin{bmatrix} \varepsilon_t & 0 & \cdots & 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} \sigma_\varepsilon^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdots & \cdot \\ 0 & \cdot & \cdots & 0 \end{bmatrix}, \qquad R = 0$$

9.4.3. Stochastic volatility

The closest equivalent to an AR or ARMA process in the second moment is probably the stochastic volatility family of models, not ARCH or GARCH. Stochastic volatility can be appropriately represented by the unobserved state variable. Unlike the previous two cases that can be and are usually estimated using traditional time series methods, such as that of the Box-Jenkins, stochastic volatility models are tested in the state space with the Kalman filter as a superior and feasible way of execution.

Define a simple time-vary variance process as:

$$y_t = \omega_t$$
$$\omega_t = \sigma_t \varepsilon_t \tag{9.26}$$
$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

In a stochastic volatility model, $h_t = \log \sigma_t^2$, the logarithm of the variance, behaves exactly as a stochastic process in the mean, such as random walks or autoregression.

$$h_t = c + \lambda h_{t-1} + \zeta_t$$
$$\zeta_t \sim N(0, \sigma_\zeta^2) \tag{9.27}$$

Equation (9.26) can be expressed as:

$$g_t = h_t + \kappa_t \tag{9.28}$$

where $g_t = \ln(y_t^2)$, and $\kappa_t = \ln(\varepsilon_t^2)$.

The observation equation is:

$$g_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} h_t \\ h_{t-1} \end{bmatrix} + \kappa_t \tag{9.29}$$

The state equation is:

$$\begin{bmatrix} h_{t+1} \\ h_t \end{bmatrix} = \begin{bmatrix} \lambda & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} h_t \\ h_{t-1} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} c + \begin{bmatrix} \zeta_{t+1} \\ 0 \end{bmatrix} \tag{9.30}$$

The construction elements are:

$$y_t = g_t, \qquad \xi_t' = \begin{bmatrix} h_t & h_{t-1} \end{bmatrix}, \qquad x_t = c$$

$$H = \begin{bmatrix} 1 & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} \lambda & 0 \\ 1 & 0 \end{bmatrix}, \qquad A = 0, \qquad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\mu_t = \kappa_t, \qquad v_t' = \begin{bmatrix} \zeta_t & 0 \end{bmatrix}$$

$$Q = \begin{bmatrix} \sigma_\zeta^2 & 0 \\ 0 & 0 \end{bmatrix}, \qquad R = \sigma_\kappa^2$$

As the model is not log normal (i.e., y_t^2 is not log normal or $\mathrm{Ln}(y_t^2)$ is not normal), it cannot be estimated by the usual maximum likelihood method. Nevertheless, when the random variables in concern are orthogonal, maximising the likelihood function will yield exactly the same estimates of the parameters, except for the standard errors of the parameters, which can be calculated by a different formula. This procedure is referred to as the Quasi Maximum Likelihood (QML) method, suggested by White (1982). Specifically, the QML estimation of stochastic volatility models is discussed in Harvey *et al*. (1994) and Ruiz (1994). Other estimation procedures include the Monte Carlo Maximum Likelihood suggested by Sandmann and Koopman (1998), where the basic stochastic volatility model is expressed as a linear state space model with log χ^2 disturbances. The likelihood function is approximated by decomposing it into a Gaussian part, estimated by the Kalman filter, and the rest is evaluated by simulation.

9.4.4. Time-varying coefficients

Specify a simple market model modified by using time-varying coefficients:

$$R_t = \alpha_t + \beta_t R_{mt} + \varepsilon_t$$
$$\alpha_t = \alpha_{t-1} + v_{1t} \tag{9.31}$$
$$\beta_t = \beta_{t-1} + v_{2t}$$

where R_t return on an individual security, R_{mt} is return on the market, and the coefficients follow a random walk.

The observation equation is:

$$R_t = \begin{bmatrix} 1 & R_{mt} \end{bmatrix} \begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} + \varepsilon_t \tag{9.32}$$

The state equation is:

$$\begin{bmatrix} \alpha_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \alpha_{t-1} \\ \beta_{t-1} \end{bmatrix} \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \tag{9.33}$$

Therefore, the construction elements of the model are:

$$y_t = R_t, \qquad \xi_t' = \begin{bmatrix} \alpha_t & \beta_t \end{bmatrix},$$

$$H(z_t) = \begin{bmatrix} 1 & R_{mt} \end{bmatrix}, \qquad F(z_t) = \begin{bmatrix} 1 & 1 \end{bmatrix}, \qquad A = 0$$

$$\mu_t = \varepsilon_t, \qquad v_t' = \begin{bmatrix} v_{1t} & v_{2t} \end{bmatrix}$$

$$Q = \begin{bmatrix} \sigma_{v1t}^2 & 0 \\ 0 & \sigma_{v2t}^2 \end{bmatrix}, \qquad R = \sigma_{\varepsilon t}^2$$

9.5. Examples and cases

Example 9.1

This is an example of decomposing the GDP series into trend and cycle components. The data used are US GDP from the first quarter of 1950 to the fourth quarter of 1999. Unlike Clark (1987), where the growth rate is a pure random walk, the model in this example has a stochastic growth rate that can be stationary or non-stationary depending on the value of λ in equation (9.36). Specifically, if λ is smaller than but close to one, the growth rate is persistent in its behaviour. The model is as follows:

$$Y_t = T_t + C_t \tag{9.34}$$

$$T_t = T_{t-1} + g_{t-1} + u_t \tag{9.35}$$

$$g_t = gc + \lambda g_{t-1} + w_t \tag{9.36}$$

$$C_t = \varphi_1 C_{t-1} + \varphi_2 C_{t-2} + v_t \tag{9.37}$$

where Y_t is log GDP; T_t is its trend component follows a random walk with a stochastic drift or growth rate which is an autoregressive process; C_t is the cycle component. Equation (9.36) collapses to the Clarke growth equation

---

when restrictions $gc = 0$ and $\lambda = 1$ are imposed. There are other reasonable assumptions. If λ is set to be zero, then the growth rate is a stationary stochastic series around a constant mean value. The growth rate is constant over time when w_t is zero as well. So, in the empirical inquiries, there are three sets of restrictions imposed against the general form of equation (9.36).

Write equations (9.34)–(9.37) in the state space form, the observation equation is:

$$
Y_t = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} T_t \\ C_t \\ C_{t-1} \\ g_t \end{bmatrix} \tag{9.38}
$$

The state equation is:

$$
\begin{bmatrix} T_{t+1} \\ C_{t+1} \\ C_t \\ g_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \varphi_1 & \varphi_2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \lambda \end{bmatrix} \begin{bmatrix} T_t \\ C_t \\ C_{t-1} \\ g_t \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ gc \end{bmatrix} + \begin{bmatrix} v_{t+1} \\ u_{t+1} \\ 0 \\ w_{t+1} \end{bmatrix} \tag{9.39}
$$

The construction elements of the model are:

$$
y_t = Y_t, \qquad \xi_t' = \begin{bmatrix} T_t & C_t & C_{t-1} & g_t \end{bmatrix}, \qquad x_t' = \begin{bmatrix} 0 & 0 & 0 & gc \end{bmatrix}
$$

$$
H = \begin{bmatrix} 1 & 1 & 0 & 0 \end{bmatrix}, \qquad F = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & \varphi_1 & \varphi_2 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \lambda \end{bmatrix}, \qquad A = 0, \quad B = 1
$$

$$
\mu_t = 0, \qquad v_t' = \begin{bmatrix} u_t & v_t & 0 & w_t \end{bmatrix}
$$

$$
Q = \begin{bmatrix} \sigma_u^2 & 0 & 0 & 0 \\ 0 & \sigma_v^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_w^2 \end{bmatrix}, \qquad R = 0
$$

The estimation results are reported in Table 9.1. Graphs of the trend, cycle, and growth rate are plotted in Figure 9.1. Inspecting the three standard deviations can give us some ideas about the behaviour of the GDP series. σ_v, the standard deviation of the cycle component, measures the contribution of cycles. There are no cyclical fluctuations when σ_v is zero ($\varphi_1 + \varphi_2$ must be zero at the same time for the cycles to be stochastic). If σ_w, the standard deviation of the growth rate, and λ are zero, the time series collapses to a

Continued

Table 9.1 Decomposition of US GDP into trend and cycle with a stochastic growth rate using the Kalman filter

φ_1	1.4978***	$(0.3203e^{-1})$
φ_2	-0.5698***	$(0.3356e^{-1})$
gc	$0.2255e^{-2}$***	$(0.1178e^{-4})$
λ	0.7362***	$(0.1892e^{-1})$
σ_u	$0.4125e^{-2}$***	$(0.8007e^{-3})$
σ_v	$0.7789e^{-2}$***	$(0.4679e^{-3})$
σ_w	$0.1175e^{-2}$**	$(0.5716e^{-3})$
Likelihood	822.3353	
LR: $gc = 0$, $\lambda = 1$	4.4006	
LR: $\lambda = 0$	0.6158	
LR: $\sigma_w = 0$	4.1664	

**significant at the 5 per cent level; ***significant at the 1 per cent level. Standard errors in parentheses.
LR is the likelihood ratio statistic, the critical values at the 10 per cent level are 2.7055 for df 1, 4.6052 for df 2 and 6.2514 for df 3.

constant growth rate case. When $\lambda < 1$ the time series is $I(1)$ and when $\lambda = 1$, i.e., a random walk growth rate is assumed, the time series is $I(2)$. The time series is a pure random walk if σ_w and σ_v are both zero while σ_u, the standard deviation of the trend, is not. Therefore, the relative importance and size of σ_u, σ_v, and σ_w, together with φ_1 and φ_2, demonstrate the behaviour of the GDP series.

It can be seen in Table 9.1 that $\varphi_1 + \varphi_2 = 0.9280$, showing a stationary cycle. The average quarterly growth rate over the period is $gc/(1 - \lambda) = 0.85$ per cent, or 3.5 per cent annually. The standard deviation of the cycle, σ_v, is nearly twice of that in the trend, σ_u; nevertheless, σ_w also contributes to the total volatility of the trend. All the estimates are significant at the 1 per cent level except for σ_w, which is also much smaller than the other two standard deviations, suggesting a stable growth rate in GDP, possibly approximated by a constant. The three figures depict the components of GDP. Figure 9.1 shows that the growth rate can swing as much as 0.2 per cent in a quarter or 0.8 per cent in one year. The growth was declining since the 1950s until the early 1980s, similar to what Clark (1987) suggested. It is most evident from the middle of the 1960s, when the US was in deep domestic crises, coupled with and highlighted by the Vietnam War. Policy changes in 1981 stimulated the economy but the prosper was proved not long lasting, due to lack of capital investment and capital formation induced by the new policy, which contributed to the Republicans' loss in a seemingly secured presidential election in 1992. The US economy has achieved

Log US GDP

Log US GDP

Log US GDP
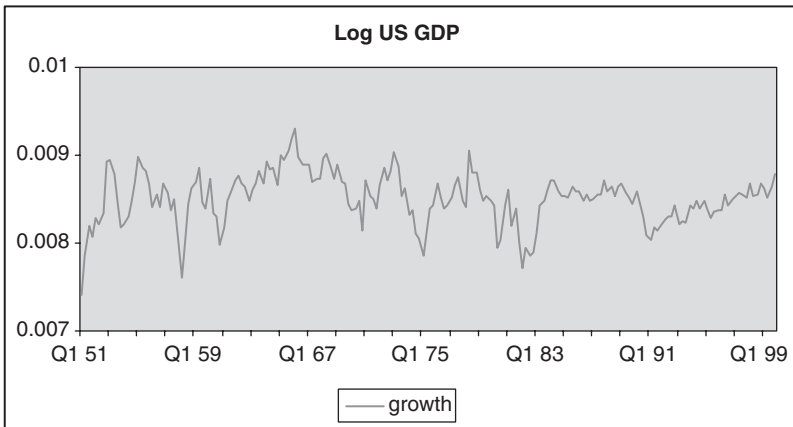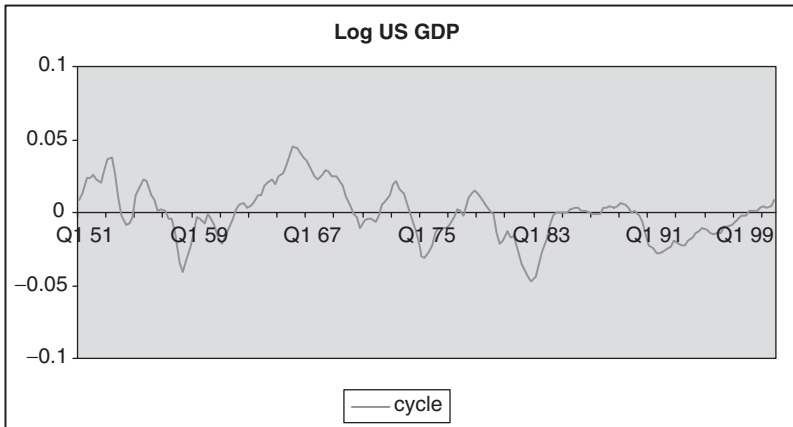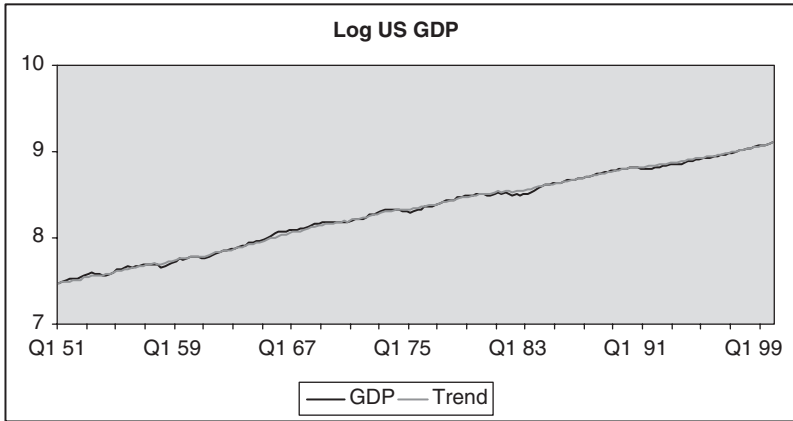
Figure 9.1 Trend, cycle and growth in US GDP.

Continued

in most of the 1990s with consistently increasing and stable growth in GDP, started from the Gulf war period and the collapse of the Soviet Union.

The likelihood ratio test does not reject any restrictions, though the random walk growth hypothesis is very close to being marginally rejected. Ranking in accordance with the likelihood function value, growth is best described as a mean-reverting stochastic process, followed by a constant plus white noise growth rate, a constant growth rate, and a random walk growth rate is the least favourable. The results show that different views on, and explanations to, some economic behaviour can be largely right at the same time.

Example 9.2

This is an example from Foresi *et al.* (1997) on interest rate models that are crucial to bond pricing. Only those parts relevant to the state space model are extracted here. The bond price is usually modelled as a function of the short-term interest rate in the bond pricing literature; and the short-term interest rate, usually called the short rate, follows some kind of generalised Wiener processes. The idea of the paper is simply that the nominal bond price is determined by the riskless real short-term interest rate and the expected instantaneous inflation rate. As the two variables are unobservable, a state space specification is proved helpful.

Basically, the paper specifies two unobserved state variables, the riskless real short-term interest rate, r_t, and the expected instantaneous inflation rate, π_t, as a vector of bivariate generalised Wiener processes (t subscript suppressed):

$$\begin{aligned}
\mathrm{d}r &= (a_1 + b_{11}r + b_{12}\pi) + \sigma_r \mathrm{d}z_r \\
\mathrm{d}\pi &= (a_2 + b_{21}r + b_{22}\pi) + \sigma_\pi \mathrm{d}z_\pi
\end{aligned} \tag{9.40}$$

Then the continuously-compounded nominal yield on zero-coupon bonds, at time t and having τ periods to maturity, $y_{n,t,\tau}$, and inflation forecast at time t for $t + \tau$, $y_{i,t,\tau}$, are treated as functions of above state variables (t subscript suppressed):

$$\begin{aligned}
y_{n,\tau} &= j_{n,\tau} + \alpha_{n,\tau,11}r + \alpha_{n,\tau,12}\pi + \varepsilon_n \\
y_{i,\tau} &= j_{i,\tau} + \alpha_{i,\tau,21}r + \alpha_{i,\tau,22}\pi + \varepsilon_i
\end{aligned} \tag{9.41}$$

where $j_{n,\tau}$ and $j_{i,\tau}$ are functions of τ and independent of the state variables, which are therefore not analysed here.

The model was estimated with the steady state instantaneous real interest rate being set to 2 per cent and 2.5 per cent respectively. The paper only reports the estimates of the state equations' results as these estimates reveal the dynamics of real interest rate and inflation processes. It illustrates the observation equations' results by plotting the term structure for both nominal bonds and indexed bonds. The relevant results are provided in the Table 9.2.

With the restricted model, b_{11} is positive but b_{12} is negative and the absolute value of b_{12} is larger, so mean reversion in the interest rate appears to be caused by the effect of the expected instantaneous inflation rate. Similarly, in the inflation equation, the parameter for the interest rate b_{21} is positive, the parameter for the inflation variable b_{22} is negative, and the absolute value of b_{22} is larger. These results suggest that the riskless real short-term interest rate is likely to push itself and the expected instantaneous inflation rate away from their steady state levels, and the expected instantaneous inflation rates tend to pull both variables back to their respective steady state levels. Notice that the interest rate has a longer half-life of 6.5 years, to

Table 9.2 US real interest rate and expected inflation processes

	Unrestricted		Restricted $b_{11} = b_{21} = 0$	
	$r^{ss} = 2.0\%$	$r^{ss} = 2.5\%$	$r^{ss} = 2.0\%$	$r^{ss} = 2.5\%$
B_{11}	0.2938***	0.2881***	−0.0344***	−0.0344***
	(0.0820)	(0.0822)	(0.0015)	(0.0015)
B_{12}	−0.4193***	−0.4273***	0	0
	(0.0971)	(0.1000)		
B_{21}	0.8240***	0.8080***	0	0
	(0.1106)	(0.1099)		
B_{22}	−1.0930***	−1.0875***	−0.7732***	−0.7733***
	(0.0828)	(0.0831)	(0.0083)	(0.0083)
σ_r	0.0100***	0.0102***	0.0151***	0.0151***
	(0.0006)	(0.0006)	(0.0005)	(0.0005)
σ_π	0.0169***	0.0168***	0.0229***	0.0229***
	(0.0007)	(0.0007)	(0.0008)	(0.0008)
$\rho_{r\pi}$	0.8235***	0.8213***	−0.1260***	−0.1263***
	(0.2414)	(0.2439)	(0.0464)	(0.0464)
Half-life r	6.54 years	6.50 years	20.17 years	20.17 years
Half-life π	1.07 years	1.06 years	0.90 years	0.90 years
Log likelihood	44.5038	44.4865	44.1516	44.1505

**significant at the 5 per cent level; ***significant at the 1 per cent level. Standard errors in parentheses.

Continued

compare with a half-life of 1.1 years for expected inflation. This reinforces the claim that there is stronger mean-reverting tendency in expected inflation than in the interest rate. When b_{12} and b_{21} are set to zero as in the restricted model, any mean reversion in a variable must come from itself. That is, b_{11} and b_{22} must be negative. The estimates in the table are negative as expected; and the size of b_{22} is much larger, reflecting that there is far stronger mean-reverting tendency in expected inflation. Any joint movement in the two variables is now through the correlation between dz_π and dz_r, as the inter-temporal links have been cut off. In the restricted model, the instantaneous correlation between the interest rate variable and the expected inflation variable is around -0.13, a number appears to be more reasonable than its counterpart in the unrestricted model, which is 0.82.

The authors claim that the Kalman filter method enables them to identify the separate influences of real rates of return and inflation expectations. Based on the Kalman filter parameter estimates, they could achieve improvements in construct yield curves and calculating investors' required premia for risk from changes in real interest rates and inflation. There is one point subject to further scrutiny: the paper says that the restricted model, where b_{12} and b_{21} are set to 0, or the interest rate process and the inflation process have no inter-temporal causal relationship and any link between them is their instantaneous correlation, performs better and the yield curves constructed from the restricted model appear to be more realistic. Then the query is: can this be justified by the economics of interest rate-inflation dynamics or is this an estimation problem technically? The half-life of 20 years for the interest rate in the restricted model also seems to be rather long. The reason could be simple: changes in the real interest rate are responses to changes in the economic environment, realised, anticipated and/or unanticipated. Without other economic variables playing a role, the evolution path of the interest rate is unlikely to change or to be altered, resulting in a longer half-life.

9.6. Empirical literature

There are growing applications of the Kalman filter in state space models, but the number is small relative to other popular models such as cointegration and GARCH. One of the reasons is that state space models are not easy to implement. On the one hand, most econometric software packages either do not have Kalman filter procedures or have the procedures which are too basic to be of practical use. On the other hand, estimates of parameters are rather sensitive to the choice of initial values and other settings of the filter.

Recent use of the Kalman filter can be found in financial markets and the economy, at micro and macro levels. In bond pricing and interest rate models, Babbs and Nowman (1998) estimate a two-factor term structure model which

allows for measurement errors by using the Kalman filter. Duan and Simonato (1999) and the above example by Foresi *et al.* (1997) are similar cases. All these studies claim that the state space model provides good fit to the yield curves of concern. Jegadeesh and Pennacchi (1996) model the target level of the interest rate, to which the short-term interest is to revert, in the state space, in a two-factor equilibrium model of the term structure. They compare the term structure of spot LIBOR and Eurodollar futures volatility to that predicted by their two-factor model and find significant improvements over the one-factor model that does not include the target level of the interest rate.

On stock market behaviour, Gallagher *et al.* (1997) decompose the stock indices of 16 countries into transitory (cycles), permanent (trends) and seasonal components. They find evidence of mean-reversion in stock prices and conclude that stock prices are not pure random walks, though the transitory component is small and does not explain more than 5 per cent of stock price variations for 12 of the 16 countries. Jochum (1999) hypothesises that the risk premium on the Swiss stock market consists of two components: the amount of volatility and the unit price of risk. The unit price of risk is time-varying and estimated by the Kalman filter, so investors' behaviour can be examined in different phases of market movement. McKenzie *et al.* (2000) estimate a time-varying beta model for the Australian market using the Kalman filter. The study is one of typical examples of time-varying coefficient models. Using the cumulative sum of squares (CUSUMSQ) test, they find beta parameter instability for all 24 industries inspected when the world market index is the relevant benchmark of the model, as the recursively estimated residuals exceeded the 5 per cent critical boundary. They find beta instability for 20 out of the 24 industry betas when the domestic market index serves as the benchmark. They conclude that time-varying betas estimated relative to the domestic index, though not universally superior, are preferred in certain circumstances. Whether the slightly inferior performance of the world index model is caused by more instability in betas has yet to be examined, though the graphs in the paper appear to suggest so.

The dividend payment pattern is one of the areas where state space models are of empirical relevance. The information content of dividends is debatably important in practice and in research, whereas it can only be inferred. In a traditional way of interpreting dividends as a long-run performance signal arising from information asymmetry, Daniels *et al.* (1997) investigate whether and how dividends are related to earnings by decomposing earnings into permanent and transitory components using the Kalman filter. They examine 30 firms' dividends and earnings and claim that there is a more robust relationship between dividends and permanent earnings, compared with that between dividends and current earnings. Marseguerra (1997) models insider information regarding the firm as an unobserved state variable that can be inferred through dividends and earnings announcements, and finds that information contained in dividend announcements varies and is a supplement to the information set already available to the market.

Other various applications include Moosa (1999) that extracts the cyclical components of unemployment and output; Serletis and King (1997) on trends and convergence of EU stock markets using a time-varying parameter model; and

Daniels and Tirtiroglu (1998) on decomposition of total factor productivity of US commercial banking into stochastic trend and cycle components.

Broadly speaking, decompose trend and cycle in economic time series, e.g., GDP, industrial production, the unemployment rate, stock prices and stock indices; unobserved variables, e.g., expectations, real interest rates, real costs; permanent and transitory components; time-vary parameters, e.g., time-varying betas.

Questions and problems

1   What is the state variable and what is an unobserved component in a state space model?
2   Discuss the advantages of the state space model and the difficulties in the empirical implementation of the model.
3   Describe the three steps of the Kalman filter algorithm in estimating a state space model.
4   Collect data from various sources, and estimate the following time series using the conventional ARIMA and in the state space using the Kalman filter (using RATS, GAUSS or other packages):

   (a) GDP of selected countries,
   (b) total return series of Tesco, Sainsbury's and ICI.

   Compare your results from the two approaches.
5   The implementation of the Kalman filter is always complicated and the results may be sensitive to even slightly different settings. To practice, collect data from various sources and repeat the same procedure of Example 1 for GDP of a few selected countries.

References

Babbs, S.H. and Nowman, K.B. (1998), Econometric analysis of a continuous time multi-factor generalized Vasicek term structure model: International evidence, *Asia-Pacific Financial Markets*, **5**(2), 59–183.

Clark, P.K. (1987), The cyclical component of the U.S. economic activity, *Quarterly Journal of Economics*, **102**(4), 797–814.

Daniels, K., Shin, T.S. and Lee, C.F. (1997), The information content of dividend hypothesis: a permanent income approach, *International Review of Economics and Finance*, **6**(1), 77–86.

Daniels, K.N. and Tirtiroglu, D. (1998), Total factor productivity growth in U.S. commercial banking for 1935–1991: A latent variable approach using the Kalman filter, *Journal of Financial Services Research*, **13**(2), 119–135.

Duan, J.C. and Simonato, J.G. (1999), Estimating and testing exponential-affine term structure models by Kalman Filter, *Review of Quantitative Finance and Accounting*, **13**(2), 111–135.

Foresi, S., Penati, A. and Pennacchi, G. (1997), Estimating the cost of U.S. indexed bonds, *Federal Reserve Bank of Cleveland Working Paper* 9701.

Gallagher, L.A. Sarno, L. and Taylor, M.P. (1997), Estimating the mean-reverting component in stock prices: A cross-country comparison, *Scottish Journal of Political Economy*, **44**(5), 566–582.

Hamilton, J.D. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press, Cambridge, England.

Harvey, A.C., Ruiz, E. and Sheppard, N. (1994), Multivariate stochastic variance models, *Review of Economic Studies*, **61**(2), 247–264.

Jegadeesh, N. and Pennacchi, G.G. (1996), The behavior of interest rates implied by the term structure of eurodollar futures, *Journal of Money, Credit and Banking*, **28**(3), 426–446.

Jochum, C. (1999), Volatility spillovers and the price of risk: Evidence from the Swiss stock market, *Empirical Economics*, **24**(2), 303–322.

Kalman, R.E. (1960), A new approach to linear filtering and prediction problems, *Transactions of ASME Series D*, *Journal of Basic Engineering*, **82**, 35–45.

Kalman, R.E. (1963), New methods of Wiener filtering theory, in Bogdanoff, J.L. and Kozin, F. (eds), *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability*, 270–388, Wiley, New York.

Kalman, R.E. (1978), A retrospective after twenty years: from the pure to the applied, *Chapman Conference on Applications of the Kalman filter to Hydrology, Hydraulics and Water Resources*, American Geophysical Union, Pittsburgh.

Kalman, R.E. and Bucy, R.S. (1961), New results in linear filtering and prediction problems, *Transactions of ASME Series D*, *Journal of Basic Engineering*, **83**, 95–108.

Marseguerra, G. (1997), The information content of dividends: an application of the Kalman filter, *Rivista Internazionale di Scienze Economiche e Commerciali*, **44**(4), 725–751.

McKenzie, M.D., Brooks, R.D. and Faff, R.W. (2000), The use of domestic and world market indexes in the estimation of time-varying betas, *Journal of Multinational Financial Management*, **10**(1), 91–106.

Moosa, I.A. (1999), Cyclical output, cyclical unemployment, and Okun's coefficient: a structural time series approach, *International Review of Economics and Finance*, **8**, 293–304.

Ruiz, E. (1994), Quasi-maximum likelihood estimation of stochastic volatility models, *Journal of Econometrics*, **63**(1), 289–306.

Sandmann, G. and Koopman, S.J. (1998), Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics*, **87**(2), 271–301.

Serletis, A. and King, M. (1997), Common stochastic trends and convergence of European Union stock markets, *The Manchester School of Economic and Social Studies*, **65**(1), 44–57.

White, H. (1982), Maximum likelihood estimation of misspecified models, *Econometrica*, **50**, 1–25.

10 Frequency domain analysis of time series

Spectral analysis, or studies in the frequency domain, is one of the unconventional subjects in time series econometrics. The frequency domain method has existed for a long time and has been extensively used in electronic engineering such as signal processing, communications, and automatic control. Although the application of the frequency domain method in econometrics may have as long a history as that in engineering, it has been sporadic and regarded unorthodox and often plays a supplementary role.

Analysis in the frequency domain does not bring in new or additional information, it is simply an alternative method with which information is observed, abstracted and processed. This is sometimes helpful. Depending on the characteristics of the issues, analysis in one domain may be more powerful than in the other. For example, cycles are better and more explicitly observed and represented in the frequency domain; while correlations in the time domain and cross spectra in the frequency domain deal with the relationship between two time series from different perspectives and, in the meantime, have defined links.

This chapter first introduces the Fourier transform, which is one of the most commonly used transformations of time series and the spectrum, the frequency domain expression of time series. In the similar spirit of covariance analysis, cross spectra, coherence and phases in multivariate time series are discussed next. In the following two sections of the chapter, frequency domain representations of commonly used time series processes are presented and frequency domain persistence measures are developed.

10.1. The Fourier transform and spectra

A continuous non-periodic time series has a continuous Fourier spectrum. For a periodic time series, its Fourier transform is discrete Fourier series. We only introduce the Fourier transform for non-periodic time series, as periodicity is rare in economic and financial time series. We do so to avoid confusion also. Then we quickly proceed to the discrete Fourier transform, which is most common in finance and economics. Let $f(t)\,(-\infty < t < \infty)$ be a continuous non-periodic time

series, then its Fourier transform (FT) is defined as:

$$F(\omega) = \int_{t=-\infty}^{\infty} f(t)\,e^{-j\omega t}\,dt \tag{10.1}$$

$F(\omega)$ is also called the spectral density function of $f(t)$. There exists an inverse Fourier transform (IFT), which is continuous, so that:

$$f(t) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(\omega)\,e^{j\omega t}\,d\omega \tag{10.2}$$

One of the most important and relevant properties of the Fourier transform is time delay or lags. Let $F(\omega)$ be the Fourier transform of $f(t)$, then the Fourier transform of $f(t - t_0)$ is $e^{-j\omega t_0}\,F(\omega)$. This can be proved briefly as follows:

$$\int_{t=-\infty}^{\infty} f(t - t_0)\,e^{-j\omega t}\,dt = \int_{t=-\infty}^{\infty} f(t)\,e^{-j\omega(t+t_0)}\,dt$$

$$= e^{-j\omega t_0} \int_{t=-\infty}^{\infty} f(t)\,e^{-j\omega t}\,dt = e^{-j\omega t_0} F(\omega)$$

In practice, for a discrete time series with N observations, such as in economics and finance, the Fourier transform would usually be the discrete Fourier transform (DFT). The pair of DFT and inverse discrete Fourier transform (IDFT) is:

$$F(k) = \sum_{n=-(N-1)}^{N-1} f(n)\,e^{-jn(2\pi k/N)} \tag{10.1}$$

and

$$f(n) = \frac{1}{N} \sum_{k=-(N-1)}^{N-1} F(k)\,e^{jk(2\pi n/N)} \tag{10.2}$$

with $\Delta\omega = 2\pi/N$, $F(k) = F(k\Delta\omega) = F(2\pi k/N)$. That is, time domain series can be expressed with different frequency components. Equation (10.1) or (10.1) is the energy spectrum. In the case of stochastic processes, the Fourier transform is concerned with the power spectrum or the power spectral density function (which can be simply called spectral density function when there is no confusion).[1]

The spectral density function of a discrete random process $\Delta_1 X_t = X_t - X_{t-1}$ $(t = 1, \dots N)$ is:

$$h(k) = \sum_{\tau=-(N-1)}^{N-1} R(\tau)\, e^{-j\tau(2\pi k/N)} \tag{10.3}$$

where $R(\tau)$ is the autocovariance function of $\Delta_1 X_t$, i.e. $R(\tau) = E\{(\Delta_1 X_t - \mu)(\Delta_1 X_{t-\tau} - \mu)\}$ and $\mu = E\{\Delta_1 X_t\}$. The inverse Fourier transform of equation (10.3) is:

$$R(\tau) = \frac{1}{N} \sum_{k=-(N-1)}^{N-1} h(k)\, e^{jk(2\pi\tau/N)} \tag{10.4}$$

Setting $\tau = 0$ in equation (10.4), we have:

$$R(0) = E\{(\Delta_1 X_t)^2\} = \frac{1}{N} \sum_{k=-(N-1)}^{N-1} h(k)\, e^{jk(2\pi\tau/N)} \tag{10.5}$$

It is the mean squared value of the process and has the meaning of power of the process, so equation (10.3) is called the power spectrum. Equation (10.1) or (10.1), in contrast, is the energy spectrum as it has the features of electrical current or voltage.

$R(\tau)$ usually takes real values and is an even function, i.e. $R(-\tau) = R(\tau)$. Accordingly, the spectral density function can be written as:

$$h(k) = \sigma_X^2 + 2 \sum_{\tau=1}^{N-1} R(\tau) \cos\left(\frac{2\pi\tau k}{N}\right) \tag{10.6}$$

Empirically, $h(k)$ has to be truncated and estimated. The simplest way of truncation is to let $R(\tau)$ pass through a rectangular window or 'truncated periodogram' window, i.e.:

$$\hat{h}(k) = \sum_{\tau=-M}^{M} R(\tau) \cos\left(\frac{2\pi\tau k}{N}\right) \tag{10.7}$$

So:

$$E\left\{\hat{h}(k)\right\} = \sum_{\tau=-M}^{M} \left(1 - \frac{|\tau|}{N}\right) R(\tau)\, \cos\left(\frac{2\pi\tau k}{N}\right) \to h(k), \quad \text{as } M \to \infty \tag{10.8}$$

In general, the truncated spectral density function takes the form:

$$\hat{h}(k) = \sum_{\tau=-M}^{M} l(\tau)R(\tau) \cos\left(\frac{2\pi\tau k}{N}\right) \tag{10.9}$$

where $l(\tau)$ is the window function. The variance of $\hat{h}(k)$ is:

$$\text{Var}\left[\hat{h}(k)\right] \sim (1+\delta_{k,0})h^2(k)\frac{1}{N}\left\{\sum_{\tau=-(N-1)}^{N-1} l_N^2(\tau)\right\} \tag{10.10}$$

$$= (1+\delta_{k,0})h^2(k)\frac{1}{N}\sum_{\theta=(N-1)}^{N-1} W_N^2(\theta)$$

where

$$W_N(\theta) = \sum_{\theta=-(N-1)}^{N-1} l_N^2(\tau)\, e^{-j\theta(2\pi k/N)} \tag{10.11}$$

is the spectral expression of the window, and $\delta_{k,0}$ is impulse function taking value of unity at $k=0$.

A rectangular window, though simple, does not perform well due to its sudden change at the cut-off points which may produce some peculiar frequency components. The Bartlett window is usually used. It is defined as:

$$l(\tau) = \begin{cases} 1 - \dfrac{|\tau|}{M}, & |\tau| \le M \\ 0 & |\tau| > M \end{cases} \tag{10.12}$$

With Bartlett's window, the variance of $\hat{h}(k)$ is:

$$\text{Var}\left[\hat{h}(k)\right] \sim \frac{2M}{3N}h^2(k), \quad \text{for } k \ne 0 \tag{10.13}$$

and

$$\text{Var}\left[\hat{h}(k)\right] \sim \frac{4M}{3N}h^2(k), \quad \text{for } k = 0 \tag{10.14}$$

If k takes value of zero in equation (10.8) it becomes:

$$E\left\{\hat{h}(0)\right\} = \sum_{\tau=-M}^{M} \left(1 - \frac{|\tau|}{N}\right)R(\tau)$$

$$= R(0) + 2\sum_{\tau=1}^{M} \left(1 - \frac{|\tau|}{N}\right)R(\tau) \tag{10.15}$$

Equation (10.15) is, in fact, the M period variance of $\Delta_M X_t = X_t - X_{t-M}$. Dividing equation (10.15) by the variance of $\Delta_1 X_t$, $\sigma^2_{\Delta_1 X}$, yields Cochrane's (1988) version of persistence. Therefore, the Cochrane measure is a specific case of equation (10.8), and assesses the long-run behaviour of time series at the zero frequency only. It appears that such measures, as represented by Campbell and Mankiw (1987a, b), Cochrane (1988) and Pesaran *et al.* (1993) are the necessary condition(s) for a random walk, not sufficient conditions, as other points on the spectrum are not evaluated against the random walk hypothesis and it is possible that they deviate from unity jointly significantly. There are no significance test statistics associated with them either, though $\mathrm{Var}\big[\hat{h}(0)\big]$ is available to provide a guideline for the accuracy of the measures, which is decided by the ratio of the window size to the number of observations, M/N, only. In other words, the window size should be small relative to the number of observations to achieve reliability in the measure. To investigate persistence and associated time series properties properly, the whole spectrum of the time series should be examined, instead of at the zero frequency point. These will be proposed and conducted in the following section.

10.2. Multivariate spectra, phases and coherence

If we replace $R(\tau)$, the autocovariance function of $\Delta_1 X_t$, by the covariance between two time series, i.e. $\mathrm{Cov}_{X,Y}(\tau) = E\{(\Delta_1 X_t - \mu_X)(\Delta_1 Y_{t-\tau} - \mu_Y)\}$, $\mu_X = E\{\Delta_1 X_t\}$ and $\mu_Y = E\{\Delta_1 Y_t\}$, then we get the cross spectrum of the two time series in the form of:

$$h_{X,Y}(k) = \sum_{\tau=-(N-1)}^{N-1} \mathrm{Cov}_{X,Y}(\tau)\, e^{-j\tau(2\pi k/N)} \tag{10.16}$$

$\mathrm{Cov}(\tau)$ is in general not an even function, so equation (10.16) cannot take the form of equation (10.6), and $h_{X,Y}(k)$ is in general a complex number:

$$h_{X,Y}(k) = c(k)\,\cos\left(\frac{2\pi k}{N}\tau\right) + jq(k)\,\sin\left(\frac{2\pi k}{N}\tau\right) \tag{10.17}$$

Unlike the univariate Fourier transform where the imaginary part is zero, the cross spectrum has both magnitude and phase as follows:

$$m(k) = \sqrt{c^2(k) + q^2(k)} \tag{10.18}$$

and:

$$p(k) = \tan^{-1}\frac{q(k)}{c(k)} \tag{10.19}$$

Equations (10.18) and (10.19) are called magnitude spectrum and phase spectrum, respectively. It can be seen, from the above analysis, that if $\mathrm{Cov}_{X,Y}(\tau)$ is an even

function, then the phase spectrum is zero, i.e. there is no overall lead of series X_t over series Y_t, and vice versa. With equations (10.18) and (10.19), the cross spectrum can also be expressed as:

$$h_{X,Y}(k) = m(k)\, e^{jp(k)} \tag{10.20}$$

so that both magnitude and phase are shown explicitly.

Another measure of the closeness of two time series is coherence, defined, in a very similar way to the correlation coefficient, as:

$$\text{Coh}_{X,Y}(k) = \frac{h_{X,Y}(k)}{h_{X,X}^{1/2}(k) h_{Y,Y}^{1/2}(k)} \tag{10.21}$$

If we make comparison of the measures in the frequency domain with those in the time domain, then the cross spectrum of equation (10.17) is corresponding to covariance in the time domain, which is not standardised; the coherence as with equation (10.21) is corresponding to correlation in the time domain, which are standardised by the square roots of the two time series' spectra and the two time series' standard deviations respectively; and the phase of equation (10.19) addresses leads and lags. As with the non-standardised cross spectrum, the closeness of two time series is not straightforwardly observed, the measure of coherence, together with the phase measure, is widely adopted in economic and financial research.

To generalise the above bi-variate analysis to the multivariate case, let:

$$\mathbf{\Sigma} = \begin{bmatrix} \text{Cov}_{11}(\tau) & \cdots & \text{Cov}_{1m}(\tau) \\ & \cdots & \\ \text{Cov}_{m1}(\tau) & \cdots & \text{Cov}_{mm}(\tau) \end{bmatrix} \tag{10.22}$$

be the covariance matrix of an m-variable system of time series. Then the cross spectra of the time series can also be expressed in a matrix:

$$\mathbf{H} = \begin{bmatrix} h_{11}(k) & \ldots & h_{1m}(k) \\ & \cdots & \\ h_{m1}(k) & \ldots & h_{mm}(k) \end{bmatrix} \tag{10.23}$$

where $h_{il}(k)\,(i,l = 1, \ldots m)$ takes the form of equation (10.16).

10.3. Frequency domain representations of commonly used time series processes

10.3.1. AR(p) process

$$Y_t = \rho_1 Y_{t-1} + \cdots + \rho_p Y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{10.24}$$

Re-arranging equation (10.24) as:

$$Y_t - \rho_1 Y_{t-1} - \cdots - \rho_p Y_{t-p} = \varepsilon_t \tag{10.24$'$}$$

Taking the Fourier transform and applying the property of time delaying yield:

$$F_Y(k)\left[1 - \rho_1 e^{-j(2\pi k/N)} - \cdots - \rho_p e^{-jp(2\pi/N)}\right] = F_\varepsilon(k) \tag{10.25}$$

So that the power spectrum or simply the spectrum of an AR process is:

$$h_Y(k)$$

$$= \frac{\sigma_\varepsilon^2}{\left[1 - \rho_1 e^{-j(2\pi k/N)} - \cdots - \rho_p e^{-jp(2\pi k/N)}\right]\left[1 - \rho_1 e^{j(2\pi k/N)} - \cdots - \rho_p e^{jp(2\pi k/N)}\right]} \tag{10.26}$$

10.3.2. MA(q) process

$$Y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \tag{10.27}$$

The Fourier transform of this process is:

$$F_Y(k) = F_\varepsilon(k)\left[1 + \theta_1 e^{-j(2\pi k/N)} + \cdots + \theta_q e^{-jq(2\pi k/N)}\right] \tag{10.28}$$

So the spectrum is:

$$\begin{aligned} h_Y(k) &= \sigma_\varepsilon^2 \left[1 + \theta_1 e^{-j(2\pi k/N)} + \cdots + \theta_q e^{-jq(2\pi k/N)}\right] \\ &\quad \times \left[1 + \theta_1 e^{j(2\pi k/N)} + \cdots + \theta_q e^{jq(2\pi k/N)}\right] \end{aligned} \tag{10.29}$$

10.3.3. VAR (p) process

$$\mathbf{Y}_t = \mathbf{A}_1 \mathbf{Y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{Y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(0, \boldsymbol{\Sigma}) \tag{10.30}$$

where $\mathbf{Y}_t$ is an $m \times 1$ vector of variables, and $\mathbf{A}_i$, $i = 1, \ldots p$ are $m \times m$ matrices of coefficients. Taking the Fourier transform yields:

$$\mathbf{F}_Y(\mathbf{k})\left[1 - \mathbf{A}_1 e^{-j(2\pi k/N)} - \cdots - \mathbf{A}_p e^{-jp(2\pi k/N)}\right] = \mathbf{F}_\Sigma(\mathbf{k}) \tag{10.31}$$

Therefore, the spectra of the VAR process is:

$$\begin{aligned} \mathbf{h}_Y(\mathbf{k}) &= \left[1 - \mathbf{A}_1 e^{-j(2\pi k/N)} - \cdots - \mathbf{A}_p e^{-jp(2\pi k/N)}\right]^{-1} \\ &\quad \times \boldsymbol{\Sigma}\left[1 - \mathbf{A}_1 e^{j(2\pi k/N)} - \cdots - \mathbf{A}_p e^{jp(2\pi k/N)}\right]^{-1} \end{aligned} \tag{10.32}$$

10.4. Frequency domain analysis of the patterns of violation of white noise conditions

10.4.1. Statistical distributions in the frequency domain of near white noise processes

Assuming a time series $X(t)$ possesses the usual properties that it is stationary, is continuous in mean square, and has higher moments up to the fourth moment, then the spectrum of the process, or the spectral distribution function, exists with the following relationships:

$$f(\omega) = \frac{1}{2\pi N} \sum_{\tau=-N}^{N} R(\tau) \, e^{-i\tau\omega} = \frac{1}{2\pi N} \sum_{\tau=-N}^{N} R(\tau) \cos(\tau\omega)$$

$$= \frac{\sigma_X^2}{2\pi} + \frac{1}{\pi} \sum_{\tau=1}^{N} R(\tau) \cos(\tau\omega) = \frac{C_0}{2\pi N} + \frac{1}{\pi N} \sum_{\tau=1}^{N-1} C_\tau \cos(\tau\omega)$$

(10.33)

$$R(\tau) = \sigma_x^2 \int_{-\pi}^{\pi} e^{i\tau\omega} dF(\omega)$$

(10.34)

where:

$$C_\tau = \sum_{t=1}^{N-\tau} X_t X_{t+\tau}, \quad C_0 = \sum_{t=1}^{N} X_t X_t = N\sigma_X^2, \quad \text{and} \quad F(\omega) = \int_0^\omega f(\omega) d\omega$$

is the integral spectrum of the process.

For a pure white noise process, C_0 obeys a χ^2-distribution with $E\{C_0\} = N$, $\mathrm{Var}\{C_0\} = 2N$; and C_τ obey normal distributions with $E\{C_\tau\} = 0$, $\mathrm{Var}\{C_\tau\} = N$, for $\tau = 1, \dots N - 1$. In the following, we show how a white noise process is distributed in the frequency domain, and the conditions on which a particular process can be accepted as a white noise process. We call such a process near white noise processes in contrast to a pure theoretical white noise. It can be shown that:

$$\lim_{N \to \infty} P \left\{ \max_{0 \le \omega \le \pi} N^{1/2} \left| F(\omega) - \frac{\omega}{2\pi} \right| \le \alpha \right\} = P \left\{ \max_{0 \le \omega \le \pi} |\xi(\omega)| \le \alpha \right\}$$

(10.35)

where $\xi(\omega)$ is a Gaussian process with:

$$P\{\xi(0) = 0\} = 1 \tag{10.36a}$$

$$P\{\xi(\pi) = 0\} = 1 \tag{10.36b}$$

$$E\{\xi(\omega)\} = 0, \quad 0 \le \omega \le \pi \tag{10.36c}$$

$$E\{\xi(v)\xi(\omega)\} = \frac{3v(\pi - \omega)}{4\pi^2}, \quad 0 \le v < \omega \le \pi \tag{10.36d}$$

$$E\{[\xi(\omega)]^2\} = \frac{3\omega(\pi - \omega)}{4\pi^2}, \quad 0 \le \omega \le \pi \tag{10.36e}$$

Equation (10.35) can be verified as follows. The integral spectrum of the time series process, or the integral of the spectrum of equation (10.33), is:

$$F(\omega) = \int_0^\omega f(\omega)d\omega = \frac{v_n^*(\omega)}{2\pi N} + \sum_{\tau=1}^{N-1} \frac{C_\tau}{\pi N} \frac{\sin \tau \omega}{\tau} \tag{10.37}$$

where $v_n^*(\omega) = \int_0^\pi C_0(\omega)d\omega$. Then:

$$N^{1/2}\left[F_p(\omega) - \frac{\omega}{2\pi}\right] = \frac{N^{1/2}}{2\pi}\left[\frac{v_n^*(\omega)}{N} - \omega\right] + \sum_{\tau=1}^{N-1} \frac{C_\tau}{\pi N^{1/2}} \frac{\sin \tau \omega}{\tau} \tag{10.38}$$

Previous studies[2] treat C_0 as a stochastic variable in the time domain, but a constant in the frequency domain irrespective of the value of ω. There is a problem mainly concerning the boundary condition: C_0 can be greater or smaller than N at point $\omega = \pi$, which does not guarantee $F(\pi) = \frac{1}{2}[F(\pi)$ is half of the total power], the requirement that the power of the standardised time series is unity (the second term on the right-hand side of equation (10.38) is zero at point $\omega = \pi$). We resort to the Kolmogorov-Smirnov theorem for a realistic representation of the distribution of C_0.

The distribution of the first term on the right-hand side can be obtained by applying the Kolmogorov-Smirnov theorem. Define:

$$z_n(t) = N^{1/2}\left[\frac{v_n(t)}{N} - t\right], \quad 0 \le t \le 1 \tag{10.39}$$

where $v_n(t) - v_n(s)$ is the number of successes in N independent trials, with probability $t - s$ of successes in each trial; $P\{v_n(0) = 0\} = 1, P\{v_n(1) = N\} = 1;$ $E\{v_n(t) - v_n(s)\} = N(t - s), E\{[v_n(t) - v_n(s)]^2\} = 2N(t - s)[1 - (t - s)], 0 \le s < t \le 1; E\{[v_n(t_1) - v_n(s_1)][v_n(t_2) - v_n(s_2)]\} = 0, 0 \le s_1 < t_1 \le s_2 < t_2 \le 1.$ Then, the m-variable distribution of the random variables $z_n(t_1), \ldots, z_n(t_m), 0 \le t_1 \cdots < t_m \le 1$ is Gaussian, and:

$$P\{z_n(0) = 0\} = 1, \quad P\{z_n(1) = 0\} = 1 \tag{10.40a}$$

$$E\{z_n(t)\} = 0, \quad 0 \le t \le 1 \tag{10.40b}$$

$$E\{[z_n(t) - z_n(s)]^2\} = (t - s)[1 - (t - s)], \quad 0 \le s < t \le 1 \tag{10.40c}$$

Now, let $\omega = \pi t$ and define:

$$z_n^*(\omega) = \frac{N^{1/2}}{2\pi} \left[\frac{v_n^*(\omega)}{N} - \omega \right] \tag{10.41}$$

we have:

$$P\{v_n^*(0) = 0\} = 1, \qquad P\{v_n^*(\pi) = N\} = 1 \tag{10.42a}$$

$$E\{v_n^*(\omega) - v_n^*(v)\} = N(\omega - v), \quad 0 \le v < \omega \le \pi \tag{10.42b}$$

$$E\left\{[v_n^*(\omega) - v_n^*(v)]^2\right\} = N(\omega - v)[\pi - (\omega - v)], \quad 0 \le v < \omega \le \pi \tag{10.42c}$$

and, the m-variable distribution of the random variables $z_n^*(\omega_1), \dots, z_n^*(\omega_m)$, $0 \le \omega_1 < \cdots < \omega_m \le \pi$ is Gaussian, with:

$$P\{z_n^*(0) = 0\} = 1, \qquad P\{z_n^*(\pi) = 0\} = 1 \tag{10.43a}$$

$$E\{z_n^*(\omega)\} = 0, \quad 0 \le \omega \le \pi \tag{10.43b}$$

$$E\left\{[z_n^*(\omega) - z_n^*(v)]^2\right\} = \frac{(\omega - v)[\pi - (\omega - v)]}{4\pi^2}, \quad 0 \le v < \omega \le \pi \tag{10.43c}$$

Equations (10.43a)–(10.43c) imply:

$$E\left\{[z_n^*(\omega)]^2\right\} = \frac{\omega(\pi - \omega)}{4\pi^2} \tag{10.44a}$$

$$E\{z_n^*(v)z_n^*(\omega)\} = \frac{v(\pi - \omega)}{4\pi^2}, \quad 0 \le v < \omega \le \pi \tag{10.44b}$$

Equation (10.44) indicates that the first term on the right-hand side of equation (10.38) is in fact $z_n^*(\omega)$.

Let us now consider the distribution of the second term on the right-hand side. It can be observed, when N approaches infinite, that:

$$E\{s(\omega)s(v)\} = \lim_{N\to\infty} E\left\{ \left(\sum_{\tau=1}^{N-1} \frac{C_\tau}{\pi N^{1/2}} \frac{\sin \tau \omega}{\tau} \right) \left(\sum_{\tau=1}^{N-1} \frac{C_\tau}{\pi N^{1/2}} \frac{\sin \tau v}{\tau} \right) \right\}$$

$$= \lim_{N\to\infty} \frac{1}{\pi^2} \sum_{\tau=1}^{N-1} \frac{\sin \tau \omega}{\tau} \frac{\sin \tau v}{\tau} = \frac{v(\pi - \omega)}{2\pi^2}, \quad 0 \le v < \omega \le \pi \tag{10.45}$$

$$E\left\{[s(\omega)]^2\right\} = \lim_{N\to\infty} E\left\{ \left(\sum_{\tau=1}^{N-1} \frac{C_\tau}{\pi N^{1/2}} \frac{\sin \tau \omega}{\tau} \right)^2 \right\} \lim_{N\to\infty} \frac{1}{\pi^2} \sum_{\tau=1}^{N-1} \left(\frac{\sin \tau \omega}{\tau} \right)^2$$

$$= \frac{\omega(\pi - \omega)}{2\pi^2} \tag{10.46}$$

Therefore:

$$s(\omega) = \lim_{N \to \infty} \sum_{\tau=1}^{N-1} \frac{C_\tau}{\pi N^{1/2}} \frac{\sin \tau \omega}{\tau} = 2z_n^*(\omega) \tag{10.47}$$

Finally, bringing the results into equation (10.38) yields:

$$\xi(\omega) = 3z_n^*(\omega) \tag{10.48}$$

The above has verified equation (10.35).

10.4.2. *Patterns of violation of white noise conditions*

This section discusses and abstracts typical patterns in time series where white noise conditions are violated. Behaviour of a particular process will be examined, in accordance with its frequency domain characteristics, which is of more empirical relevance. From Equations (10.35) and (10.36), three patterns can be characterised with regard to their violation of white noise conditions, setting against the benchmark of a pure white noise process.

Pattern 1

Lower frequency components stochastically dominate higher frequency components in the frequency range (ϖ_1, ϖ_2) if $\xi(\omega) > 0$, $\varpi_1 < \omega < \varpi_2$. Lower frequency components stochastically *consistently* dominate higher frequency components if $\xi(\omega) > 0$, $0 < \omega < \pi$, and the time series is said to possess the features of the compounding effect.

By definition and according to theorem 1, $\xi(\omega)$ is the difference between the integral of the process under examination and the integral of a pure white noise process, being scaled by N, when $N \to \infty$, i.e.:

$$\xi(\omega) = \lim_{N \to \infty} N^{1/2} \int_0^\omega \left(I_p(\omega) - \frac{1}{2\pi} \right) d\omega$$

Figure 10.1 shows the features of such stochastic processes. The top panel of the figure is the time domain response to a unit size shock of a time series with compounding features, against a random walk response. The dashed line indicates the evolution path of the time series if no shocks have ever occurred. The middle panel is a typical spectrum for such time series, and the bottom panel is the $\xi(\omega)$ statistic for such time series. The spectrum in Figure 10.1 is a monotonous decrease function of ω, with $\xi(\omega) > 0$, $0 < \omega < \pi$, and $\xi_\omega''(\omega) < 0$, $0 < \omega < \pi$, where $\xi_\omega''(\omega)$ is the second-order derivative of $\xi(\omega)$ with respect to ω.[3] Stochastically
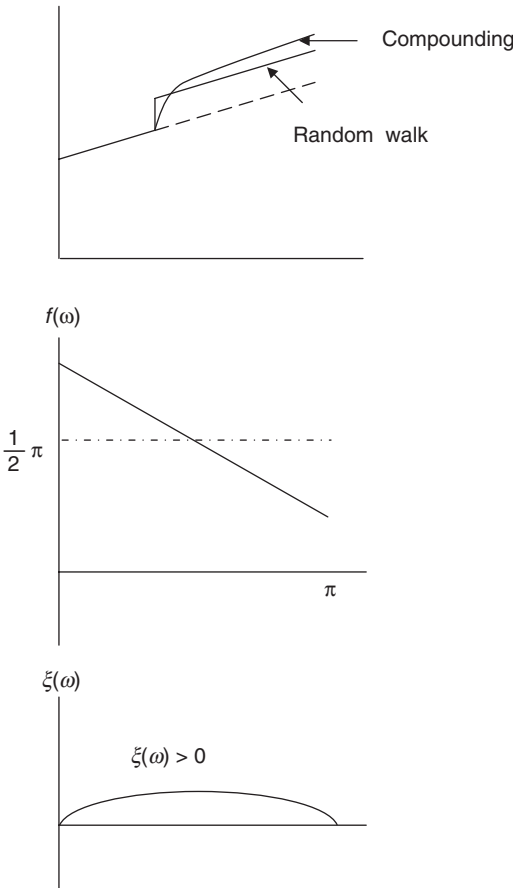
Figure 10.1 Lower frequencies dominate (compounding effect).

consistent dominance has a looser requirement than a spectrum of monotonous function.

Pattern 2

Higher frequency components stochastically dominate lower frequency components in the frequency range (ϖ_1, ϖ_2) if $\xi(\omega) < 0$, $\varpi_1 < \omega < \varpi_2$. Higher frequency components stochastically *consistently* dominate lower frequency components if $\xi(\omega) < 0$, $0 < \omega < \pi$, and the time series is said to possess mean-reverting tendencies.

Figure 10.2 shows the features of such stochastic processes. The top panel of the figure is the time domain response to a unit size shock of a time series with mean-reverting tendencies, against a random walk response. The dashed line indicates
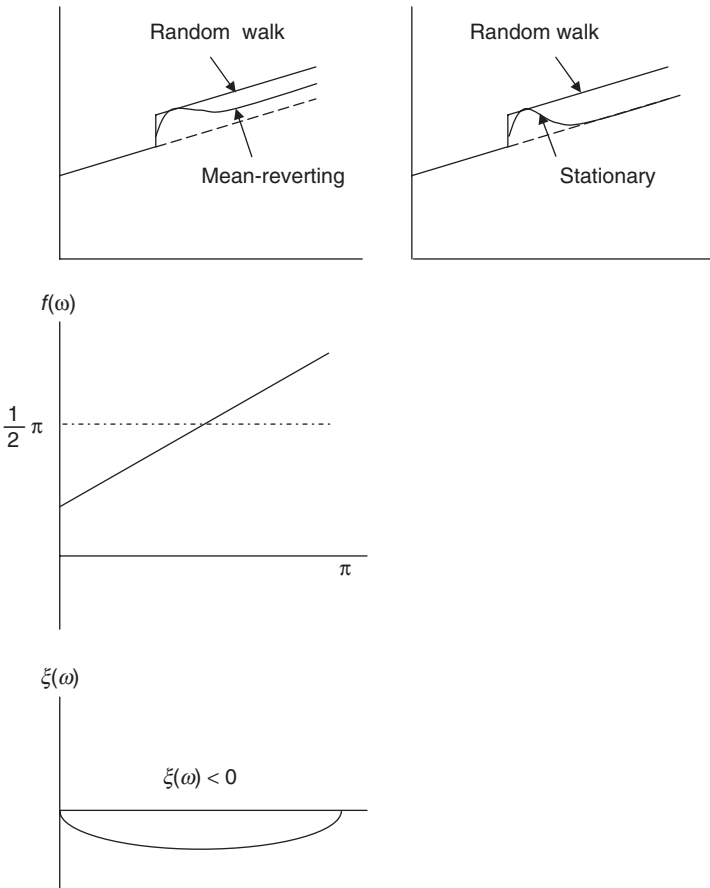
Figure 10.2 Higher frequencies dominate (mean-reverting tendency).

the evolution path when there are no shocks to the time series. The middle panel is a typical spectrum for such time series, and the bottom panel is the $\xi(\omega)$ statistic for such time series. The spectrum in Figure 10.2 is a monotonous increase function of ω, with $\xi(\omega) < 0$, $0 < \omega < \pi$, and $\xi_\omega''(\omega) > 0$, $0 < \omega < \pi$. Stochastically consistent dominance has a looser requirement than a spectrum of monotonous function. Relevant discussions for pattern 1 apply here.

Pattern 3

Higher (lower) frequency components do not stochastically *consistently* dominate lower (higher) frequency components if there exist sub-sets of frequencies ω^+, ω^- and ω^0 that $\xi(\omega) > 0$, $\omega \in \omega^+$, $\xi(\omega) < 0$, $\omega \in \omega^-$ and $\xi(\omega) = 0, \omega \in \omega^0$; and the time series is said to possess the features of mixed complexity.
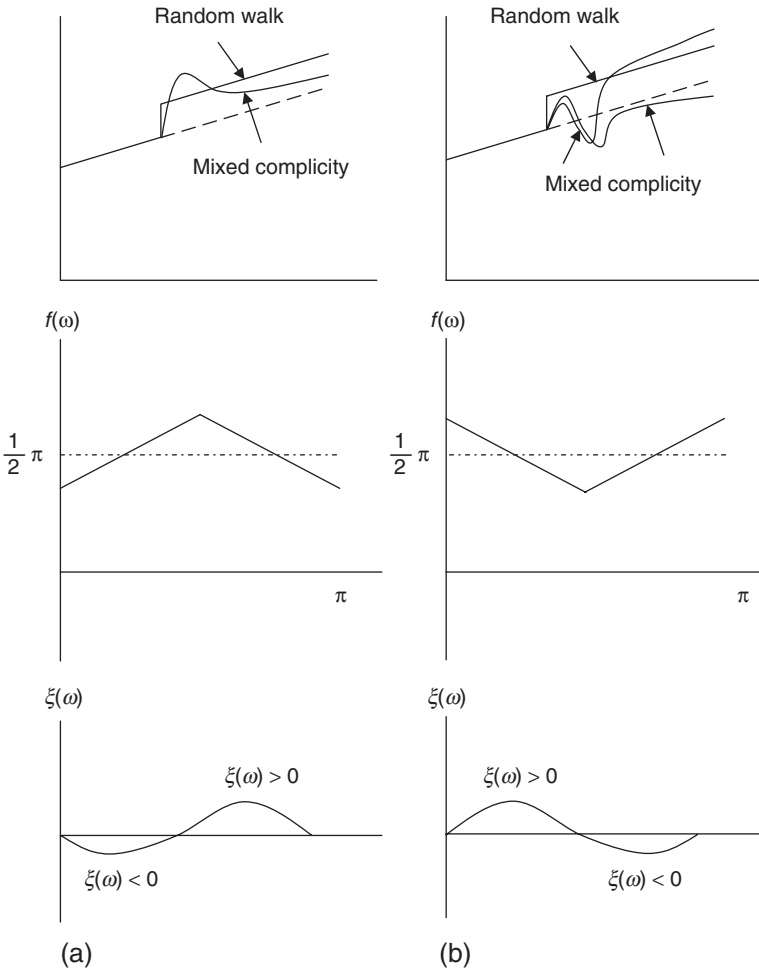
Figure 10.3 Two cases of mixed complicity.

Figure 10.3 demonstrates the features of such stochastic processes. Relevant discussions for pattern 1 apply here. Figure 10.3(a) shows a case where there are more powers in the medium range frequencies, while Figure 10.3(b) shows a case where there are more powers in the low and high frequencies. The top panel of the figures is the time domain response to a unit size shock of a time series with the features of mixed complexity, against a random walk response. The dashed line indicates the evolution path when there are no shocks to the time series. The middle panel is typical spectra for such time series, and the bottom panel is the $\xi(\omega)$ statistics for such time series.

10.5. Examples and cases

> ### Example 10.1
>
> To demonstrate the frequency domain analysis of the patterns of violation
> of white noise conditions developed in this chapter, we scrutinise business
> cycle patterns in UK sectoral output empirically. The data sets used in
> this study are UK aggregate GDP and output in seven main GDP sectors,
> starting in the first quarter, 1955, ending in the first quarter, 2002, and they
> are seasonally adjusted at the 1995 constant price. The data sets for the
> two sub-sectors within the Services sector start from the first quarter in
> 1983. The seven main sectors used in the study are: Agriculture, Forestry
> and Fishing (A&B); Manufacturing (D); Electricity, Gas and Water Supply
> (E); Construction (F); Distribution, Hotels, Catering and Repairs (G&H);
> Transport, Storage and Communication (I); and Services (J–Q, including
> business services and finance, and government and other services). The
> Mining and Quarrying sector (C) is excluded, as its weight in UK GDP
> is minimal and has been declining over decades; and more importantly,
> its change has been mainly influenced by unconventional economic forces
> and other factors. The Services sector is examined in two parts of Business
> Services and Finance (J&K) and Government and Other Services (L–Q),
> since the attributes and features of these two types of services are rather
> different and, consequently, may possess different response patterns in
> business cycle fluctuations. However, the two disaggregate services series
> only came into existence in the first quarter of 1983, instead of the first
> quarter of 1955 for the seven main sectors.
>
> The time domain summary statistics of these sectors' output and GDP
> are provided in Table 10.1. Sector J&K, Business Services and Finance,
> sector I, Transport, Storage and Communication, and sector E, Electricity,
> Gas and Water Supply, enjoy a greater than average growth rate, though
> the Business Services and Finance sector has experienced a decrease in its
> growth rate. The lowest growing sectors are A&B, Agriculture, Forestry
> and Fishing, and D, Manufacturing. The Manufacturing sector has also
> gone through a decline in its growth during this period, along with sector F,
> Construction. As being analysed above, sector L–Q, Government and Other
> Services, has the most smoothed growth, with its standard deviation in
> growth being the smallest and much smaller than that for all the other sectors.
>
> *Table 10.1* Time domain summary statistics of sectoral output and GDP
>
> | | A&B | D | E | F | G&H | I | J&K | L–Q | J–Q | GDP |
> |------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
> | Mean | 0.3810 | 0.3298 | 0.7864 | 0.4682 | 0.5732 | 0.7903 | 0.8945 | 0.4673 | 0.6188 | 0.6001 |
> | Std | 2.3046 | 1.7676 | 4.2544 | 2.7480 | 1.4162 | 1.5019 | 0.9018 | 0.3678 | 0.7042 | 1.0121 |
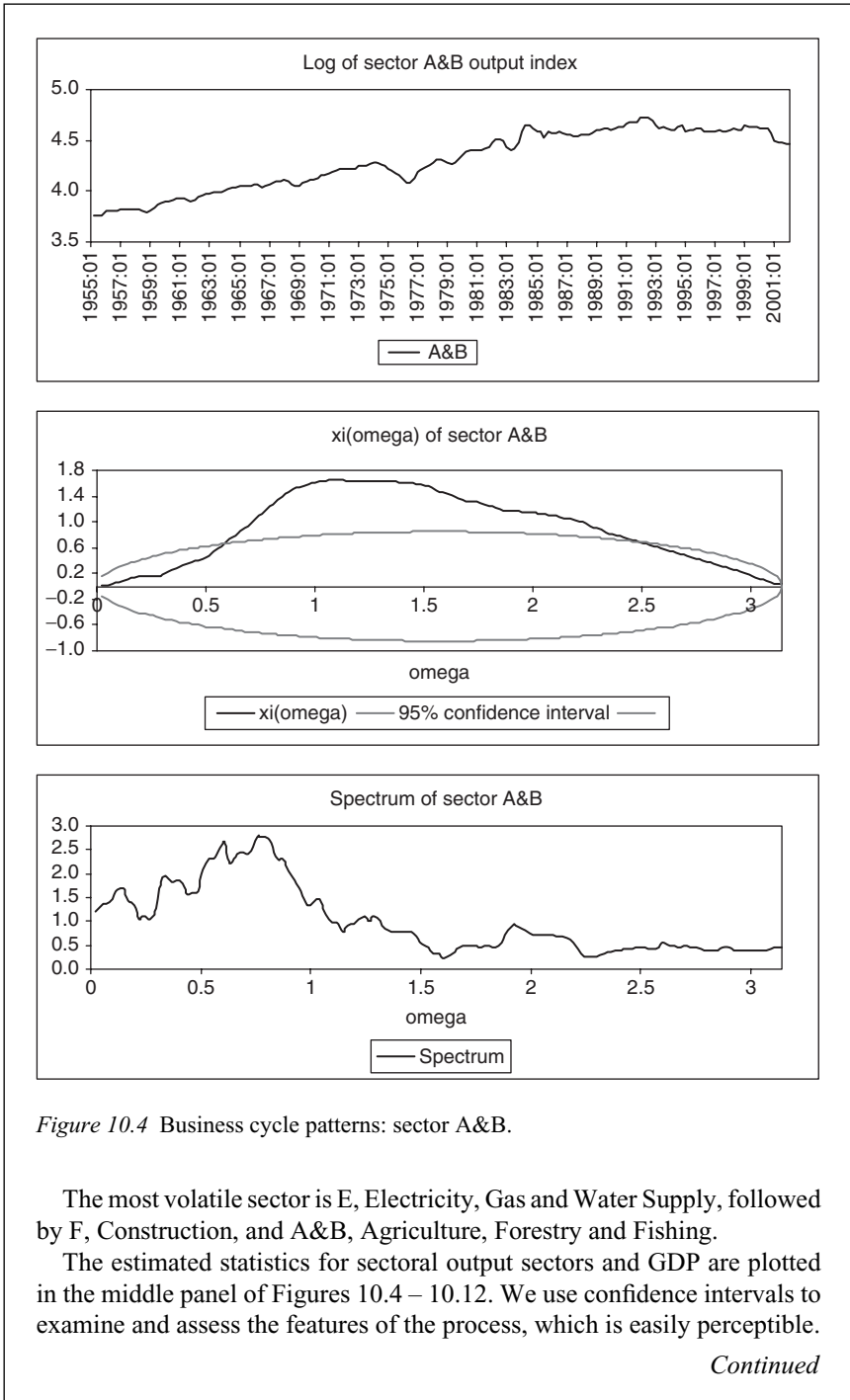
Figure 10.4 Business cycle patterns: sector A&B.

The most volatile sector is E, Electricity, Gas and Water Supply, followed by F, Construction, and A&B, Agriculture, Forestry and Fishing.

The estimated statistics for sectoral output sectors and GDP are plotted in the middle panel of Figures 10.4 – 10.12. We use confidence intervals to examine and assess the features of the process, which is easily perceptible.

Continued

Log of sector D output index

xi(omega) of sector D

Spectrum of sector D
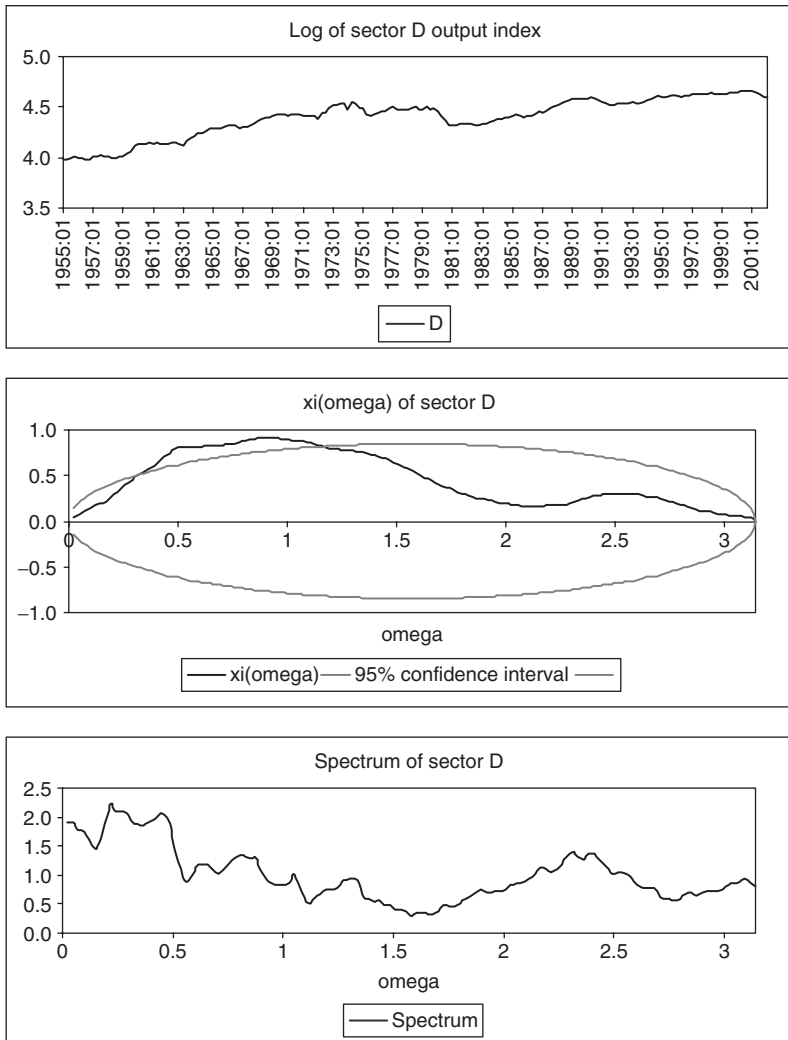
Figure 10.5 Business cycle patterns: sector D.

In addition, output series themselves are exhibited in the top panel and spectra are presented in the bottom panel of these figures. We examine the $\xi(\omega)$ statistic and inspect the associated patterns for GDP sectors in relation to their institutional features. Four sectors show the features of compounding effects to varied degrees. They are sector A&B, Agriculture, Forestry and Fishing; sector D, Manufacturing; sector I, Transport, Storage and Communication; and sector J&K, Business Services and Finance.

Log of sector E output index

xi(omega) of sector E

Spectrum of sector E
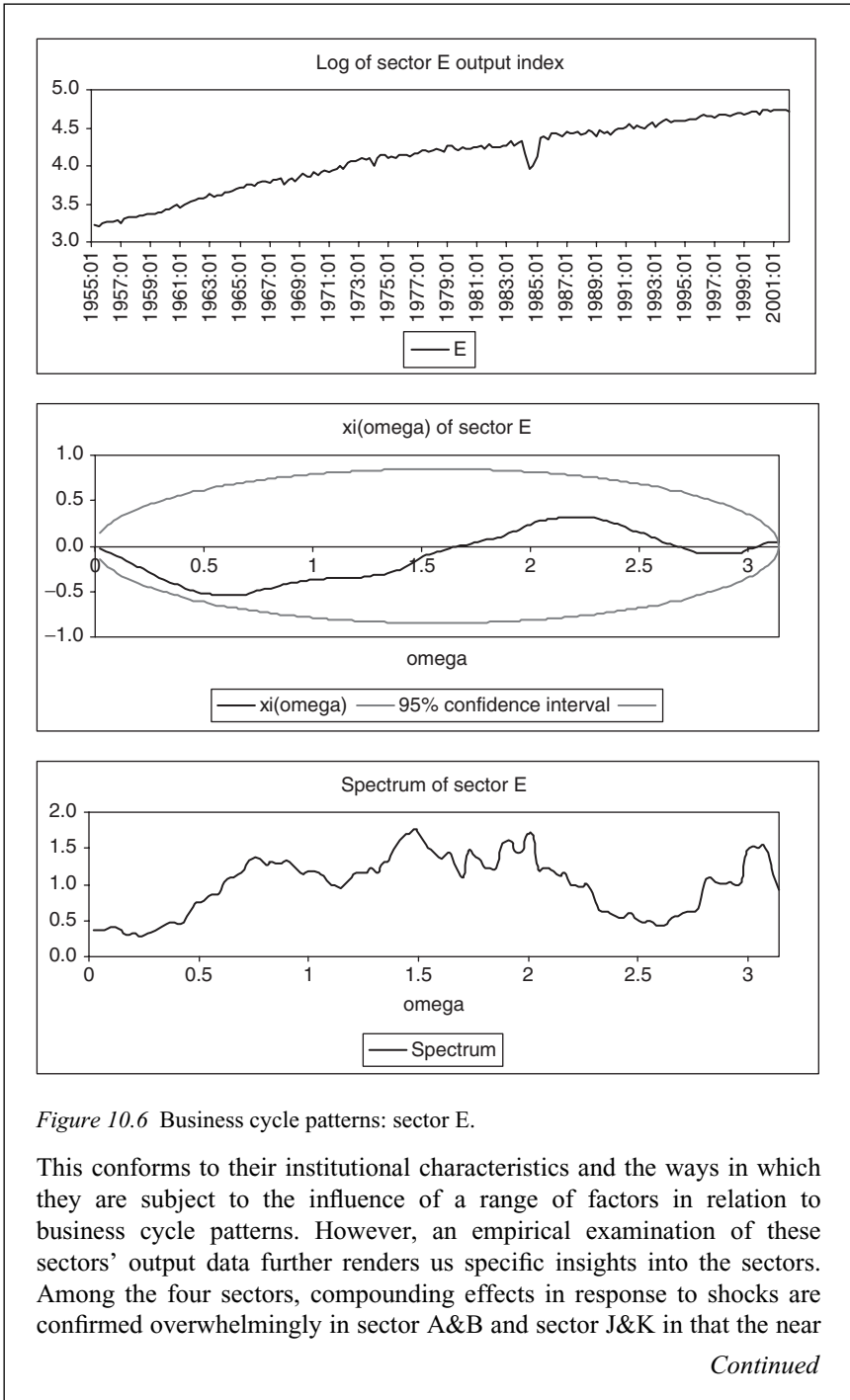
Figure 10.6 Business cycle patterns: sector E.

This conforms to their institutional characteristics and the ways in which they are subject to the influence of a range of factors in relation to business cycle patterns. However, an empirical examination of these sectors' output data further renders us specific insights into the sectors. Among the four sectors, compounding effects in response to shocks are confirmed overwhelmingly in sector A&B and sector J&K in that the near

Continued

Log of sector F output index

xi(omega) of sector F

Spectrum of sector F
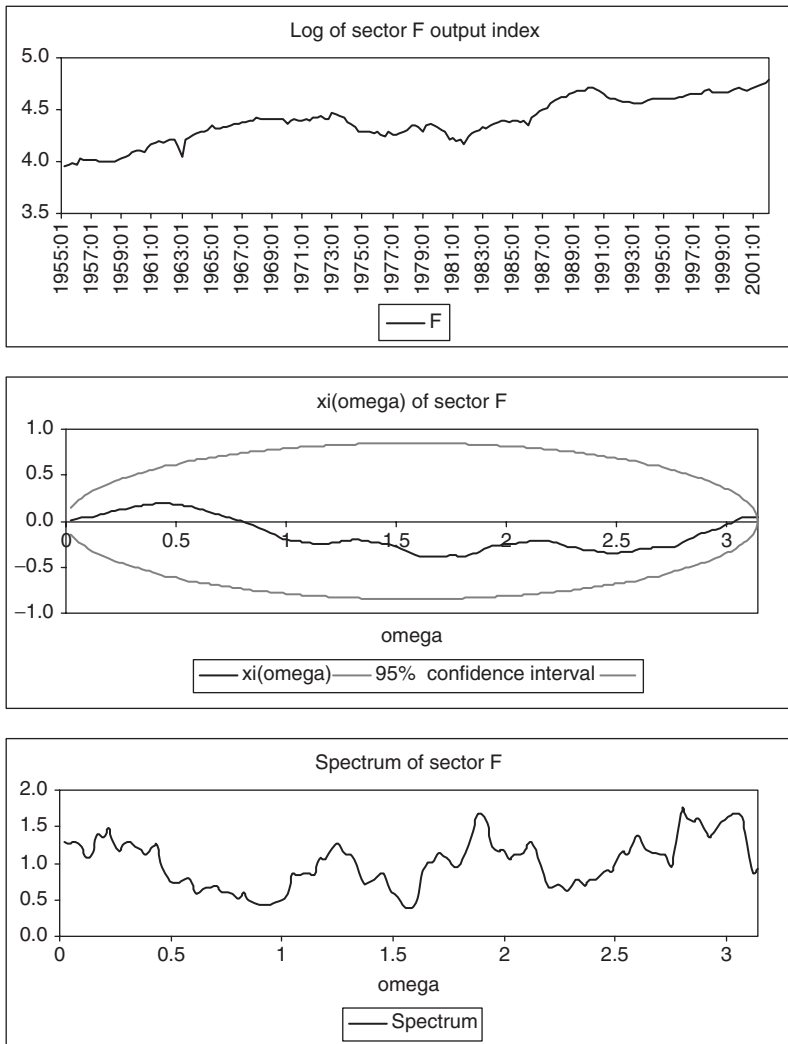
Figure 10.7 Business cycle patterns: sector F.

white noise conditions are significantly violated – as shown in Figure 10.4 and Figure 10.10, $\xi(\omega)$ statistics are positive in the whole frequency range and the majority of $\xi(\omega)$ are substantially above the upper band of the 95 per cent confidence interval. In the case of sector D, $\xi(\omega)$ statistics are positive in the whole frequency range but only a small part of $\xi(\omega)$ are beyond the upper band of the 95 per cent confidence interval, revealed by Figure 10.5. For sector I, it is observed in Figure 10.9 that most of $\xi(\omega)$

Log of sector G&H output index

xi(omega) of sector G&H

omega

xi(omega) ——— 95% confidence interval ———

Spectrum of sector G&H

omega

Spectrum
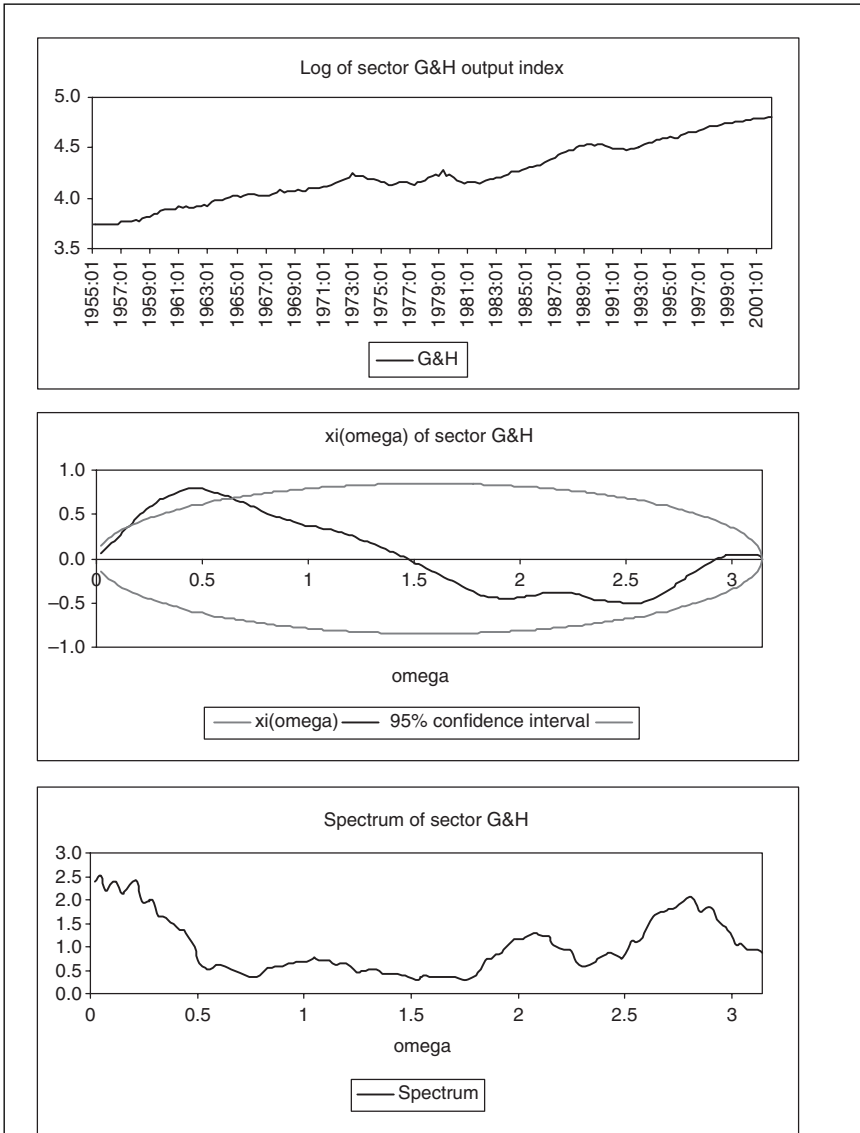
Figure 10.8 Business cycle patterns: sector G&H.

are positive and only a small part of $\xi(\omega)$ are beyond the upper band of the 95 per cent confidence interval. So, compounding effects are not as strong in sector D and sector I as in sector A&B and sector J&K. Since these sectors possess the features of compounding effects in their response

Continued
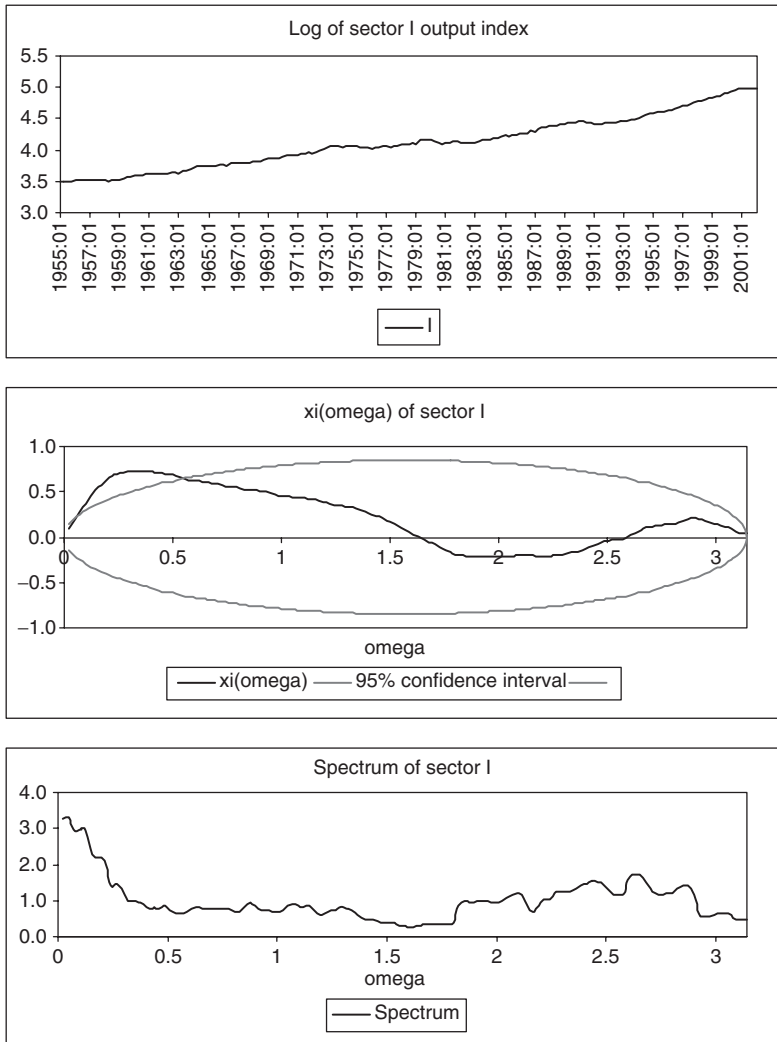
Figure 10.9 Business cycle patterns: sector I.

to shocks in business cycles, the consequence of good as well as bad events or incidents, policy related or technology based, would accumulate in the course to affect the performance of these sectors, with the Agriculture, Forestry and Fishing sector and the Business Services and Finance sector being hit the most.

   Sector E, Electricity, Gas and Water Supply, and sector F, Construction, demonstrate random walk-like behaviour – it is observed in Figure 10.6 and

Log of sector J&K output index

xi(omega) of sector J&K

Spectrum of sector J&K
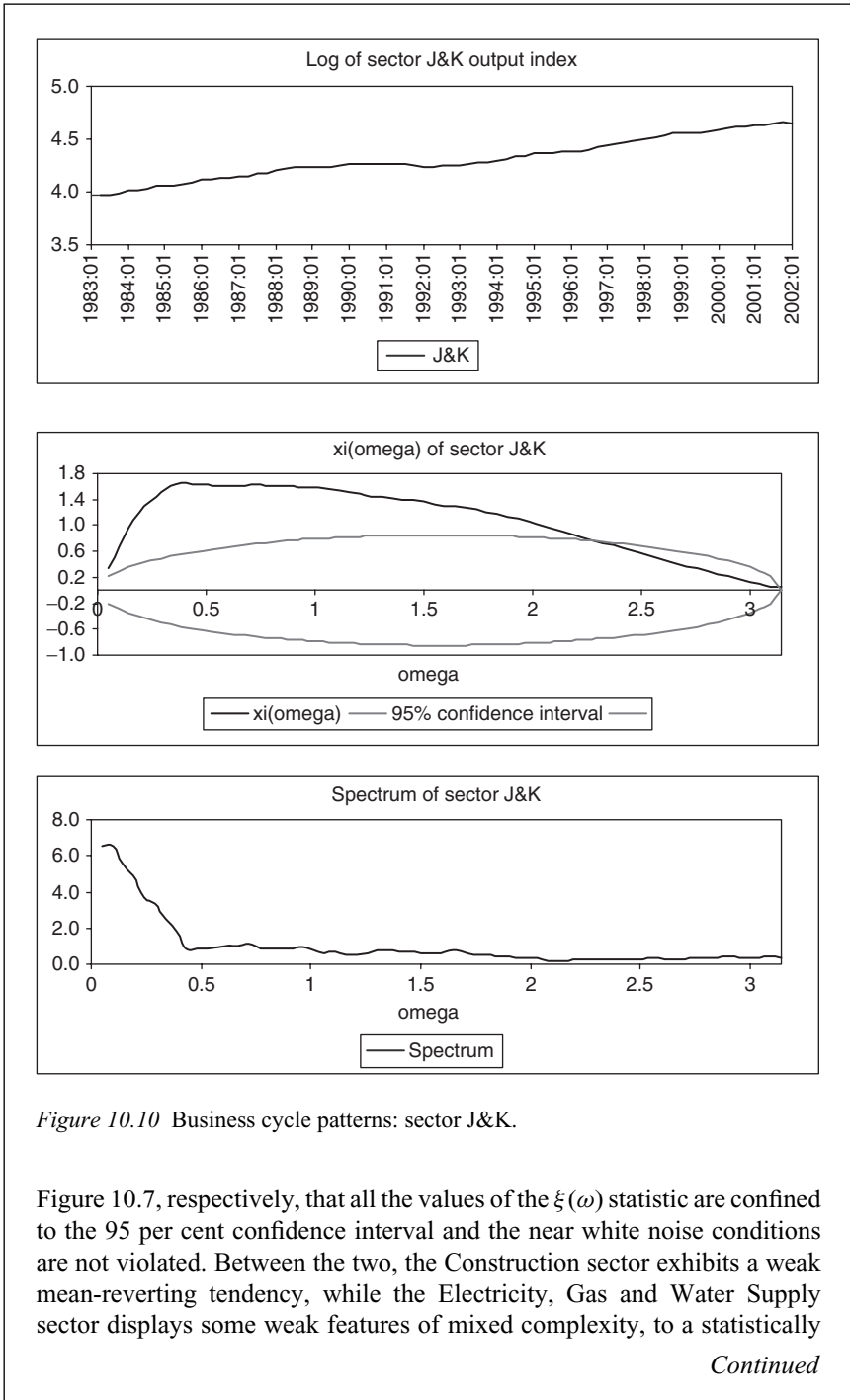
Figure 10.10 Business cycle patterns: sector J&K.

Figure 10.7, respectively, that all the values of the $\xi(\omega)$ statistic are confined to the 95 per cent confidence interval and the near white noise conditions are not violated. Between the two, the Construction sector exhibits a weak mean-reverting tendency, while the Electricity, Gas and Water Supply sector displays some weak features of mixed complexity, to a statistically

Continued

Log of sector L–Q output index

xi(omega) of sector L–Q in difference

Spectrum of sector L–Q in difference
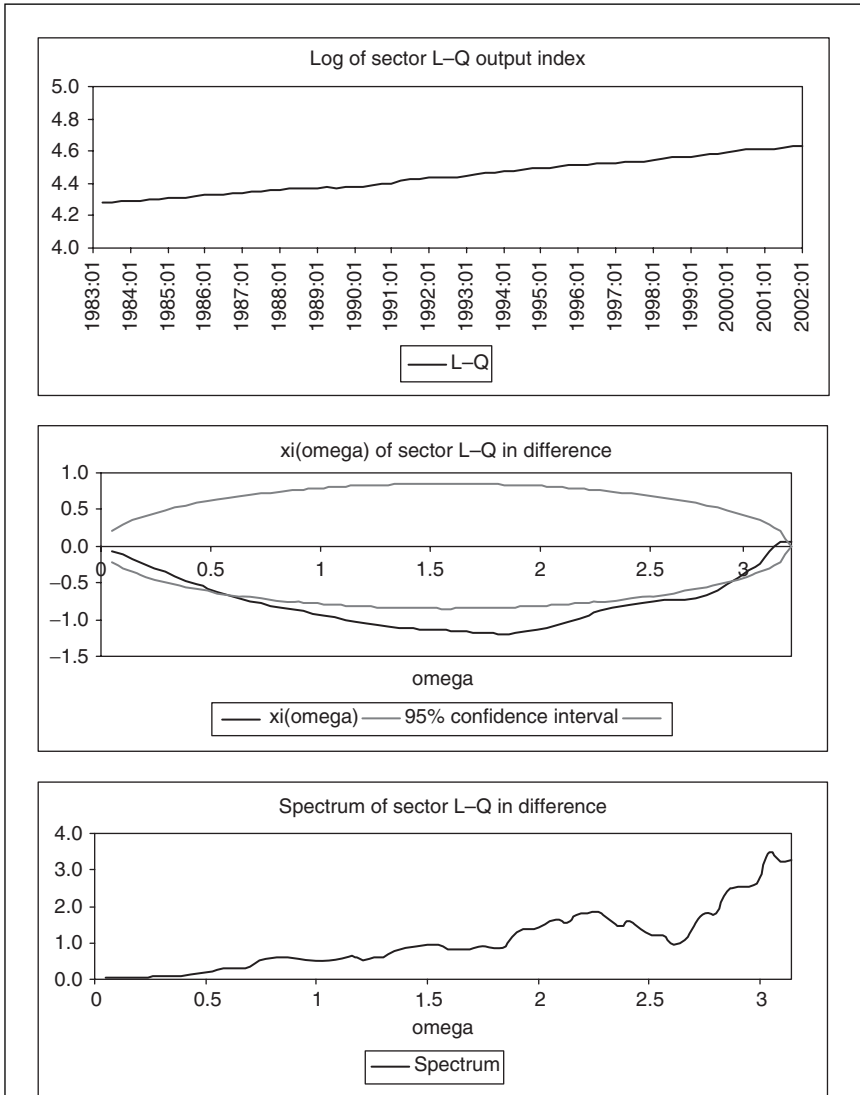
Figure 10.11 Business cycle patterns: sector L–Q.

insignificant degree. These findings also conform to the two sectors'
institutional features and indicate that, between the two sectors, the
Construction sector would display relatively less persistent response patterns
in business cycles due to its lower regulatory requirements. Sector
G&H, Distribution, Hotels, Catering and Repairs, is associated with a
mixed complicity response pattern in business cycles and exhibits some

Log of GDP

xi(omega) of GDP

Spectrum of GDP
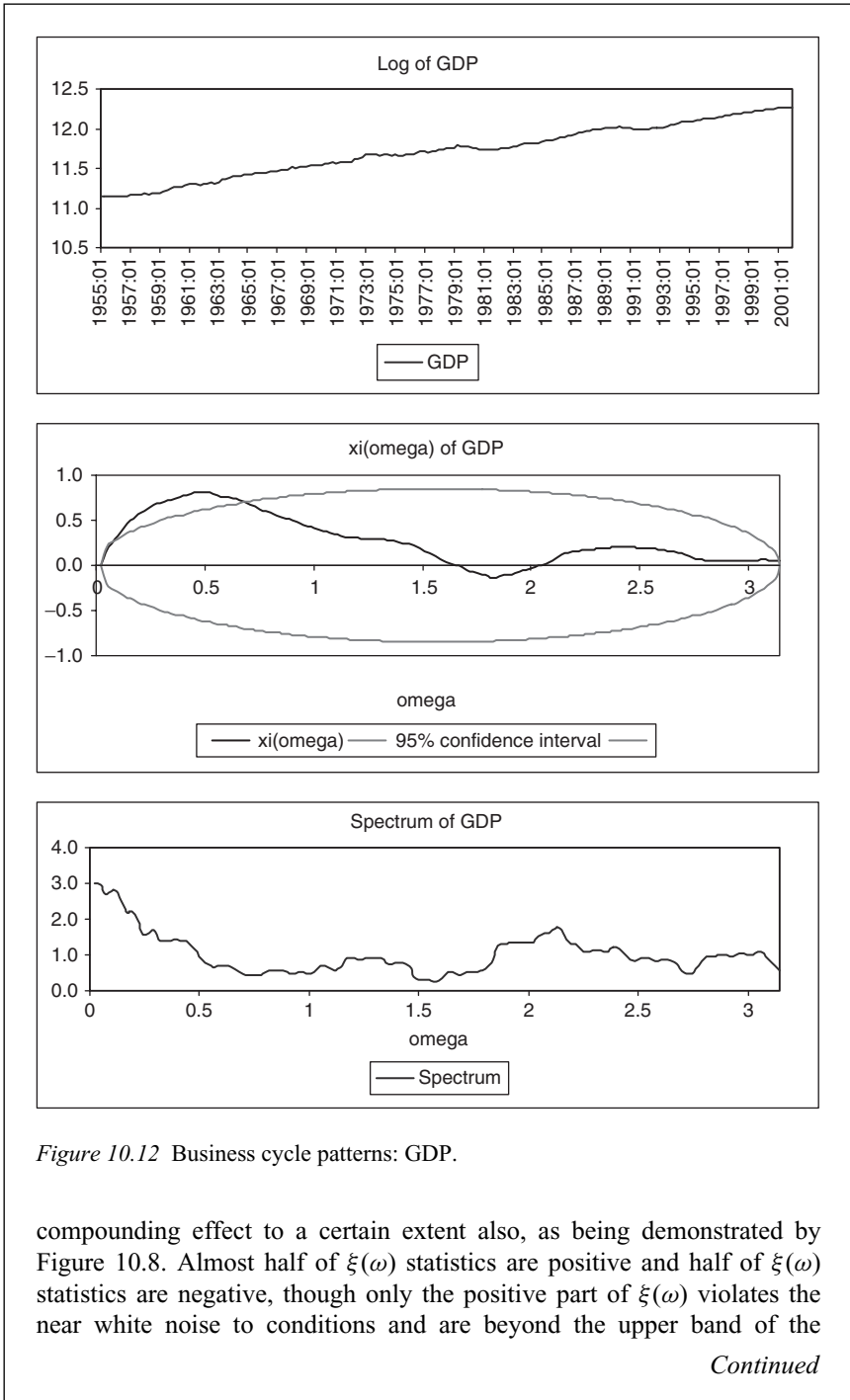
Figure 10.12 Business cycle patterns: GDP.

compounding effect to a certain extent also, as being demonstrated by Figure 10.8. Almost half of $\xi(\omega)$ statistics are positive and half of $\xi(\omega)$ statistics are negative, though only the positive part of $\xi(\omega)$ violates the near white noise to conditions and are beyond the upper band of the

Continued

95 per cent confidence interval. Some of the negative $\xi(\omega)$ statistics are close to, but yet to reach, the lower band of the 95 per cent confidence interval. These findings fit into the institutional characteristics of the Distribution, Hotels, Catering and Repairs sector fairly appropriately.

Sector L–Q, Government and Other Services, as expected, exhibits a business cycle pattern rather different from that in all other sectors, revealed by Figure 10.11. It possesses mean-reverting tendencies to such an extent that is almost for a stationary time series. All the values of the $\xi(\omega)$ statistic are negative, most of them having violated the near white noise conditions and being below the lower band of the 95 per cent confidence interval. It can be observed in Table 10.1 that the sector has the most smoothed growth, with its standard deviation in growth being much smaller than that for all all the other sectors, mainly arising from the sector's characteristics of experiencing infrequent shocks in business cycles. Smoothed growth, or a small standard deviation in growth, does not necessarily mean a lower persistence or close to being stationary. It is infrequent shocks that, to a large extent, contribute to the features demonstrated by the Government and Other Services sector.

The behaviour of the aggregate GDP must reflect the business cycle features demonstrated by GDP sectors that are dominated by persistent, sizeable compounding effects in their response to shocks in business cycles. It is observed in Figure 10.12 that the majority of $\xi(\omega)$ statistics are positive, with a few of them being beyond the upper band of the 95 per cent confidence interval or having violated the near white noise conditions. Although the result from the analysis of the aggregate GDP makes known its business cycle response patterns and features, which match the outcome and conclusion of sectoral analysis, it is sectoral analysis, in reference to the institutional background and characteristics of the sectors, that reveals how different sectors behave differently in business cycles and exhibit a specific and different business cycle pattern, and lays theoretical cornerstones for GDP's overall business cycle features.

Example 10.2

This is a case studying comovements among financial markets by means of cross spectra and phases in the frequency domain, in a paper entitled 'Pre- and post-1987 crash frequency domain analysis among Pacific rim equity markets' by Smith (2001). The paper examines five stock markets of Australia, Hong Kong, Japan, the US, and Canada pair-wise, using the individual stock market index data of Morgan Stanley International Capital Perspectives, measured in local currencies. The period surrounding the

crash, i.e. May 1987 through March 1988, is excluded from the sample and the author claims that this is due to the volatility during this period. Therefore, the pre-crash sample is from 18 August 1980 to 29 May 1987, and the post-crash period from 8 March 1988 to 16 December 1994.

Since the Fourier transform requires that the time series are stationary, routine unit root tests are carried out applying the KPSS procedure. The purpose is to confirm that (the logarithms of) the indices in levels are non-stationary while their first differences are stationary, which has been duly achieved.

Table 10.2 reports the frequency domain statistics of coherence for the pre- and post-crash periods and the Wilcoxon Z statistic for testing the hypothesis that the pre- and post-crash coherences are drawn from the same population. It also provides the time domain statistics of correlation for comparison. The correlations for pair-wise markets in any period are rather low, so these markets would be judged against having substantial links among them. However, the frequency domain peak coherences are much higher, ranging from 0.2415 between Japan and Hong Kong to 0.5818 between Canada and Australia in the pre-crash period. Nevertheless, the mean coherences are modest, suggesting that comovements are quite different at different frequencies – they are more coherent at some frequencies and less coherent at some other frequencies. It has been shown

Table 10.2 Correlation and coherence

	USHK	USAU	CAJA	CAHK	CAAU	JAHK	JAAU	AUHK
Pre correlation	0.0689	0.0661	0.0417	0.0813	0.1177	0.1394	0.1350	0.1807
Post correlation	0.0873	0.0183	0.1715	0.1072	0.0846	0.2053	0.1676	0.1680
Pre peak coherence	0.4010	0.5093	0.3461	0.3141	0.5818	0.2415	0.3338	0.3665
Post peak coherence	0.4114	0.5215	0.3349	0.4313	0.4968	0.3553	0.3085	0.3511
Pre mean coherence	0.1860	0.2482	0.1714	0.1771	0.2747	0.1502	0.1667	0.1877
Post mean coherence	0.2250	0.2659	0.2259	0.2044	0.2371	0.2108	0.1913	0.1981
Wilcoxon Z^a	-8.16[***]	-4.25[***]	-14.8[***]	-4.83[***]	4.40[***]	-17.2[***]	-6.94[***]	-1.10

a The Wilcoxon Z statistic tests the null that the coherences for the two periods are drawn from the same population.
[*]significant at the 10 per cent level; [**]significant at the 5 per cent level; [***]significant at the 1 per cent level.

that coherences are low at high frequencies. In all the cases, except the pairs of the US and Hong Kong, Canada and Hong Kong and Japan and Hong Kong, the coherence falls while the frequency increases. For the pairs of the US and Hong Kong, Canada and Hong Kong and Japan and Hong Kong, the peak occurs at the frequency between 0.1 and 0.2 (5–10 days). The paper also presents phase diagrams for the pairs of the markets. Without a consistent pattern, the phase diagrams are mainly of practical interest.

The Wilcoxon Z statistic suggests that, in every case except the pair of Australia and Hong Kong, the pre- and post-crash coherences are statistically different, or the coherences for the two periods are drawn from different populations. Moreover, both peak and mean coherences have increased in the post-crash period as against the pre-crash period in all the cases except the pair of Canada and Australia, implying increased post-crash comovements among these markets.

10.6. Empirical literature

Frequency domain analysis is most popular in business cycle research because the research object and the method match precisely. Garcia-Ferrer and Queralt's (1998) study is typical in the frequency domain–decomposing business cycles into long, medium and short term cycles following Schumpeter's work. They claim that the frequency domain properties of the time series can be exploited to forecast business cycle turning points for countries exhibiting business cycle asymmetries. Cubadda (1999) examines common features in the frequency domain and the time domain. Understandably, the author has concluded that the serial correlation common feature is not informative for the degree and the lead–lag structure of their comovements at business cycle frequencies. Since the lead–lag relationship in the frequency domain is not an exact mapping of the serial correlation common feature in the time domain, the former (latter) does not contain all the information possessed by the latter (former), but does contain additional information not possessed by the latter (former). As being pointed out earlier, transformation does not generate extra new information, it simply provides another way of viewing and processing information, which may be more effective in certain aspects. Bjornland (2000) is, technically, on business cycle phases. The author finds that consumption and investment are consistently pro-cyclical with GDP in the time domain and the frequency domain. However, the business cycle properties of real wage and prices are not so clear-cut, depending on the de-trending methods used. Although the number is considerably less than that in the traditional time domain, there are still a few empirical studies in the area from time to time, for example, Entorf (1993) on constructing leading indicators from non-balanced sectoral business

survey data, Englund *et al.* (1992) on Swedish business cycles, Canova (1994) on business cycle turning points, and King and Rebelo (1993) on the Hodrick–Prescott filter.

As analysis in the frequency domain offers a different way of examining time series properties and patterns, it is naturally applied to issues such as unit roots, VAR and cointegrated variables. Choi and Phillips (1993) develop frequency domain tests for unit roots. Their simulation results indicate that the frequency domain tests have stable size and good power in finite samples for a variety of error-generating mechanisms. The authors conclude that the frequency domain tests have some good performance characteristics in relation to time domain procedures, although they are also susceptible to size distortion when there is negative serial correlation in the errors. Olekalns (1994) also considers frequency domain analysis as an alternative to the Dickey–Fuller test. With regard to dynamic models, error correction in continuous time is considered by Phillips (1991) in the frequency domain. Stiassny (1996) proposes a frequency domain decomposition technique for structural VAR models and argues, with an example, the benefit of adopting this technique in providing another dimension of the relationships among variables. Examining univariate impulse responses in the frequency domain, Wright (1999) estimates univariate impulse response coefficients by smoothing the periodogram and then calculating the corresponding impulse response coefficients and forms the confidence intervals of the coefficients through a frequency domain bootstrap procedure.

Other empirical studies on varied topics can be found in Cohen (1999) on analysis of government spending, Wolters (1995) on the term structure of interest rates in Germany, Koren and Stiassny (1995) on the causal relation between tax and spending in Austria, Copeland and Wang (1993) on combined use of time domain and frequency domain analyses, and Bizer and Durlauf (1990) on the positive theory of government finance, to list a few.

Notes

1 This can also be the product of the Fourier transform and its conjugate.
2 For example Bartlett (1950), Grenander and Rosenblatt (1953, 1957), and Priestley (1996).
3 $\xi'_\omega(\omega) = N^{1/2} \left[I_p(\omega) - \frac{1}{2}\pi \right]$, $\xi''_\omega(\omega) = N^{1/2} I'_p(\omega)$.

Questions and problems

1 What is spectral analysis of time series? Does spectral analysis render new or more information?
2 Discuss the advantages and disadvantages of the analysis in the frequency domain.
3 Describe the Fourier transform and the inverse Fourier transform.
4 What are phases and coherence in spectral analysis? Contrast them with leads/lags and correlation in the time domain.

5   Collect data from various sources and perform the Fourier transform for the following time series (using RATS, GAUSS or other packages):

   (a) GDP of selected countries,
   (b) total return series of selected companies,
   (c) foreign exchange rates of selected countries *vis-à-vis* the US$.

6   Collect data from DataStream and estimate phases and coherence for the following pairs of time series:

   (a) the spot and forward foreign exchange rates of the UK£ *vis-à-vis* the US$,
   (b) the spot foreign exchange rates of the UK£ and Japanese yen *vis-à-vis* the US$.

7   Collect data from various sources and estimate phases and coherence for the following pairs of time series:

   (a) GDP of the US and the Canada,
   (b) GDP and retail sales of the UK.

8   Collect data from DataStream and estimate phases and coherence for the following pairs of time series:

   (a) total returns of Tesco and Sainsbury's,
   (b) total returns of Intel and Motorola.

References

Bartlett, M.S. (1950), Periodogram analysis and continuous spectra, *Biometrika*, **37**, 1–16.

Bizer, D.S. and Durlauf, S.N. (1990), Testing the positive theory of government finance, *Journal of Monetary Economics*, **26**, 123–141.

Bjornland, H.C. (2000), Detrending methods and stylised facts of business cycles in Norway: an international comparison, *Empirical Economics*, **25**, 369–392.

Campbell, J.Y. and Mankiw, N.W. (1987a), Are output fluctuations transitory?, *Quarterly Journal of Economics*, **102**, 857–880.

Campbell, J.Y. and Mankiw, N.W. (1987b), Permanent and transitory components in macroeconomic fluctuations, *American Economic Review*, **77** (Papers and Proceedings), 111–117.

Canova, F. (1994), Detrending and turning points, *European Economic Review*, **38**, 614–623.

Choi, I. and Phillips, P.C.B. (1993), Testing for a unit root by frequency domain regression, *Journal of Econometrics*, **59**, 263–286.

Cochrane, J.H. (1988), How big is the random walk in GDP?, *Journal of Political Ecomony*, **96**, 893–920.

Cohen, D. (1999), An analysis of government spending in the frequency domain, *Board of Governors of the Federal Reserve System Finance and Economics Discussion Series*: 99/26.

Copeland, L.S. and Wang, P.J. (1993), Estimating daily seasonals in financial time series: the use of high-pass spectral filters, *Economics Letters*, **43**, 1–4.

Cubadda, G. (1999), Common serial correlation and common business cycles: a cautious note, *Empirical Economics*, **24**, 529–535.

Englund, P., Persson, T. and Svensson, L.E.O. (1992), Swedish business cycles: 1861–1988, *Journal of Monetary Economics*, **30**, 343–371.

Entorf, H. (1993), Constructing leading indicators from non-balanced sectoral business survey series, *International Journal of Forecasting*, **9**, 211–225.

Garcia-Ferrer, A. and Queralt, R.A., (1998), Using long-, medium-, and short-term trends to forecast turning points in the business cycle: some international evidence, *Studies in Nonlinear Dynamics and Econometrics*, **3**, 79–105.

Grenander, U. and Rosenblatt, M. (1953), Statistical spectral analysis arising from stationary stochastic processes, *Annals of Mathematical Statistics*, **24**, 537–558.

Grenander, U. and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*, Wiley, New York.

King, R.G. and Rebelo, S.T. (1993), Low frequency filtering and real business cycles, *Journal of Economic Dynamics and Control*, **17**, 207–331.

Koren, S. and Stiassny, A. (1995), Tax and spend or spend and tax? an empirical investigation for Austria, *Empirica*, **22**, 127–149.

Olekalns, N. (1994), Testing for unit roots in seasonally adjusted data, *Economics Letters*, **45**, 273–279.

Pesaran, M.H., Pierse, R.G. and Lee, K.C., (1993), Persistence, cointegration and aggregation: a disaggregated analysis of output fluctuations in the US economy, *Journal of Econometrics*, **56**, 67–88.

Phillips, P.C.B. (1991), Error correction and long-run equilibrium in continuous time, *Econometrica*, **59**, 967–980.

Priestley, M.B. (1996), *Sprectral Analysis and Time Seties* 9th edition (1st edition 1981), Academic Press, London.

Smith, K.L. (2001), Pre- and post-1987 crash frequency domain analysis among Pacific Rim equity markets, *Journal of Multinational Financial Management*, **11**, 69–87.

Stiassny, A. (1996), A spectral decomposition for structural VAR models, *Empirical Economics*, **21**, 535–555.

Wolters, J. (1995), On the term structure of interest rates – empirical results for Germany, *Statistical Papers*, **36**, 193–214.

Wright, J.H. (1999), Frequency domain inference for univariate impulse responses, *Economics Letters*, **63**, 269–277.

11 Limited dependent variables and discrete choice models

Firms and individuals encounter choice problems from time to time. An investment decision is indeed to make a choice; so is a savings decision – whether to save and how much to save. One particular kind of choice is binary choice. For example, a firm may choose to use financial derivatives to hedge interest rate risk, or choose not to use financial derivatives. A firm may decide to expand its business into foreign markets, or not to expand into foreign markets. Further, if the firm decides to expand its business overseas, it may acquire an existing firm in the foreign country as part of its strategic plan, or establish a new plant in the foreign country. In pension provisions, if employees are entitled to choose between a defined benefit plan and a defined contribution plan, they have to make a decision of binary choice. On the other hand, employers may decide to close the defined benefit plan to new members or continue to offer the defined benefit plan to new members. There are numerous such examples of binary choice in people's daily life, firms' financing and investment activities, managers' business dealings, and financial market operations. Binary choice can be extended to discrete choice in general, i.e. there are more than two alternatives to choose from but the values of the variable are still discrete, such as 0, 1, 2, or 1, 2, 3, 4. A simple example can be firms' choice of overseas stock exchanges for dual listings, e.g. they may choose one of New York, Tokyo, or London. Binary or discrete choice models can be generalised to refer to any cases where the dependent variable is discrete, such as discrete responses and categories.

In addition to discrete choice models where a dependent variable possesses discrete values, the values of dependent variables can also be censored or truncated. That is, the variable is not observed over its whole range. A dependent variable that is discrete, truncated or censored is a limited dependent variable. This chapter addresses discrete choice models while the next chapter deals with truncation and censoring. The chapter first presents the commonly used formulations of probit and logit for binary choice models. General discrete choice models are introduced next in the multinomial logit framework, followed by ordered probit and ordered logit. Since discrete choice models are non-linear, marginal effects are considered specifically in one section. Deviating from the preceding chapters of the book, data sets analysed in this chapter and the next are primarily cross-sectional. That is,

they are data for multiple entities, such as individuals, firms, regions or countries, observed at a single time point.

11.1. Probit and logit formulations

In a binary choice model, the dependent variable takes the value of either 1 or 0, with a probability that is a function of one or more independent variables. A binary variable is sometimes said to obey a binomial distribution; while a discrete variable taking more than two discrete values is said to possess a multinomial distribution, a topic to be studied in the next section. When the dependent variable is binary or discrete, some assumptions on linear regression estimation procedures, such as the OLS where the dependent variable is continuous, do not hold. Alternative methods have to be employed for the formulation and estimation of binary and discrete choice modelling appropriately. Probit and logit models are two commonly used models to formulate binary choice, response or categorisation. The probit model derives its name from the probit function, so does the logit model from the logistic function. Regression for estimating the probit model is referred to probit regression, and regression for estimating the logit model is commonly known as logistic regression.

The underlying process of the observed binary variable is supposed to be an unobserved latent variable. Their relationship and presentation are:

$$Y^* = \mathbf{X}\boldsymbol{\beta} + \varepsilon \tag{11.1a}$$

$$Y = \begin{cases} 1, & \text{if } Y^* \geq 0 \\ 0, & \text{if } Y^* < 0 \end{cases} \tag{11.1b}$$

The unobserved variable is a continuous function of a linear combination of a set of explanatory variables, which measures the utility of an activity. If the function value or utility is high enough, the activity will be undertaken and 1 is chosen. Otherwise, if the function value or utility is not high enough, the activity will not be undertaken and 0 is chosen.

The probit function is the inverse cumulative distribution function of the normal distribution. The probit model employs an inverse probit function, so it is a kind of the cumulative standard normal distribution. Let Y be a binary variable that takes the value of either 1 or 0. The probit model is defined as:

$$P(Y = 1 \mid \mathbf{X}\boldsymbol{\beta}) = \Phi(\mathbf{X}\boldsymbol{\beta}) \tag{11.2}$$

where $\mathbf{X}$ is a vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of the corresponding coefficients, and $\Phi(z)$ is the cumulative normal distribution. The probit model states that the probability of Y taking the value of 1 follows the cumulative standard normal distribution of a linear function of the given set of explanatory variables $\mathbf{X}$.

The logit model is based on the odds of an event taking place. The logit of a number P between 0 and 1 is defined as:

$$\text{Logit}(P) = \text{Ln}\left(\frac{P}{1-P}\right) \tag{11.3}$$

If $P = P(Y = 1 | \mathbf{X\beta})$ is the probability of an event taking place, then $P/(1-P)$ is the corresponding odds and $\text{Ln}[P/(1-P)]$ is the corresponding log odds. The logit model states that the log odds of an event taking place are a linear function of a given set of explanatory variables, i.e.:

$$\text{Ln}\left(\frac{P}{1-P}\right) = \mathbf{X\beta} \tag{11.4}$$

The probability $P = P(Y = 1 | \mathbf{X\beta})$ can be solved as:

$$P(Y = 1 | \mathbf{X\beta}) = \frac{\exp(\mathbf{X\beta})}{1 + \exp(\mathbf{X\beta})} \tag{11.5}$$

Figure 11.1 demonstrates the probabilities predicted by the probit model and the logit model. It can be observed that the two curves are similar except that the curve for the probit model is sharper. The probability of the probit model increases to a greater extent at point $\mathbf{X\beta} = 0$, with the slope being 0.3989. While the probability of the logit model increases to a smaller extent at point $\mathbf{X\beta} = 0$, with the slope being 0.25. This is also shown in Figure 11.2 by their probability densities. The probability density of the probit model is more condensed than that of the logit model. The logit model has one particular advantage and makes more sense – it
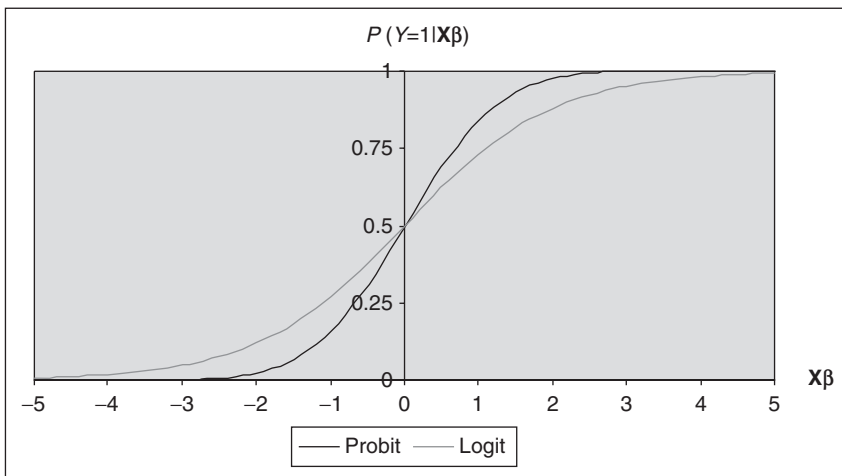
Figure 11.1 Predicted probability by probit and logit.
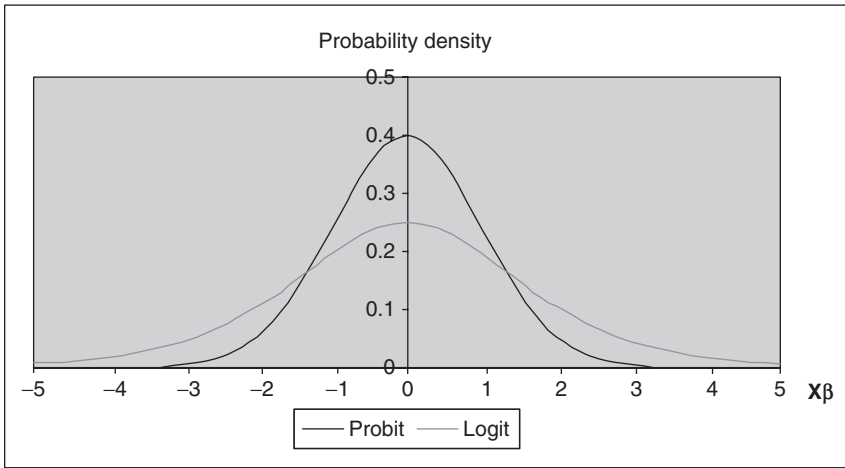
Figure 11.2 Probability density of probit and logit.

specifies that the odds are exponential and the log odds are linear functions of explanatory variables, making it easy to interpret. On the other hand, the probit model implies that the increase in the probability of the event taking place follows a standard normal distribution, in line with the assumptions on many probabilistic events.

Estimation of the probit model and the logit model is usually through maximising their likelihood function. Recall that the likelihood of a sample of N independent observations with probabilities $P_1, P_2, \ldots, P_N$, L, is:

$$L = P_1 P_2 \cdots P_N = \prod_{i=1}^{N} P_i \tag{11.6}$$

With the probit model, the probability of Y being 1 is $P(Y = 1 \mid \mathbf{X\beta}) = \Phi(\mathbf{X\beta})$ and the probability of Y being 0 is $P(Y = 0 \mid \mathbf{X\beta}) = 1 - P(Y = 1 \mid \mathbf{X\beta}) = 1 - \Phi(\mathbf{X\beta})$. Therefore, the likelihood function of the probit model, $L(\boldsymbol{\beta})$, is:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left[\Phi(\mathbf{X}_i\boldsymbol{\beta})\right]^{Y_i} \left[1 - \Phi(\mathbf{X}_i\boldsymbol{\beta})\right]^{(1-Y_i)} \tag{11.7}$$

and the log likelihood function of the probit model, $LL(\boldsymbol{\beta})$, is:

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{N} Y_i \mathrm{Ln}\left[\Phi(\mathbf{X}_i\boldsymbol{\beta})\right] + (1 - Y_i)\mathrm{Ln}\left[1 - \Phi(\mathbf{X}_i\boldsymbol{\beta})\right] \tag{11.8}$$

Similarly, for the logit model, the probability of Y being 1 is:

$$P(Y=1 \mid \mathbf{X\beta}) = \frac{\exp(\mathbf{X\beta})}{1+\exp(\mathbf{X\beta})}$$

and the probability of Y being 0 is:

$$P(Y=0 \mid \mathbf{X\beta}) = 1 - P(Y=1 \mid \mathbf{X\beta}) = 1 - \frac{\exp(\mathbf{X\beta})}{1+\exp(\mathbf{X\beta})}$$

Therefore, the likelihood function of the logit model, $L(\mathbf{\beta})$, is:

$$L(\mathbf{\beta}) = \prod_{i=1}^{N} \left[\frac{\exp(\mathbf{X}_i\mathbf{\beta})}{1+\exp(\mathbf{X}_i\mathbf{\beta})} \right]^{Y_i} \left[1 - \frac{\exp(\mathbf{X}_i\mathbf{\beta})}{1+\exp(\mathbf{X}_i\mathbf{\beta})} \right]^{(1-Y_i)} \tag{11.9}$$

and the log likelihood function of the logit model, $LL(\mathbf{\beta})$, is:

$$LL(\mathbf{\beta}) = \sum_{i=1}^{N} Y_i \mathrm{Ln} \left[\frac{\exp(\mathbf{X}_i\mathbf{\beta})}{1+\exp(\mathbf{X}_i\mathbf{\beta})} \right] + (1-Y_i)\mathrm{Ln} \left[1 - \frac{\exp(\mathbf{X}_i\mathbf{\beta})}{1+\exp(\mathbf{X}_i\mathbf{\beta})} \right]$$

$$\tag{11.10}$$

Coefficient estimates can be derived, adopting procedures that maximise the above log likelihood functions.

11.2. Multinomial logit models and multinomial logistic regression

The multinomial distribution is an extension of the binomial distribution when and where there are more than two discrete outcomes – choices, responses or categories. Estimation of multinomial logit models is usually dealt with by multinomial logistic regression. A simplest multinomial logit model is one where the dependent variable can have three discrete values of 0, 1 and 2. Let P_1 denote $P(Y=1 \mid \mathbf{X\beta}^{\alpha})$ and P_2 denote $P(Y=2 \mid \mathbf{X\gamma}^{\alpha})$. Let $P_0 = P(Y=0 \mid \mathbf{X\alpha})$ be the base category or reference category. Let $\mathbf{\beta} = \mathbf{\beta}^{\alpha} - \mathbf{\alpha}$ and $\mathbf{\gamma} = \mathbf{\gamma}^{\alpha} - \mathbf{\alpha}$. In a similar but not an exactly comparable way in which odds are applied in the previous section, we construct two binomial logit models with reference to the base category of P_0:

$$\mathrm{Ln}\left(\frac{P_1}{P_0}\right) = \mathbf{X\beta} \tag{11.11a}$$

$$\mathrm{Ln}\left(\frac{P_2}{P_0}\right) = \mathbf{X\gamma} \tag{11.11b}$$

Note $P_0 = 1 - P_1 - P_2$, so:

$$Ln\left(\frac{P_1}{P_0}\right) = Ln\left(\frac{P_1}{1 - P_1 - P_2}\right) = \mathbf{X\beta} \tag{11.12a}$$

$$Ln\left(\frac{P_2}{P_0}\right) = Ln\left(\frac{P_2}{1 - P_1 - P_2}\right) = \mathbf{X\gamma} \tag{11.12b}$$

The probability $P(Y = 1 \,|\, \mathbf{X\beta})$ can therefore be solved as:

$$P(Y = 1 \,|\, \mathbf{X\beta}) = \frac{\exp(\mathbf{X\beta})}{1 + \exp(\mathbf{X\beta}) + \exp(\mathbf{X\gamma})} \tag{11.13a}$$

the probability $P(Y = 2 \,|\, \mathbf{X\gamma})$ is:

$$P(Y = 2 \,|\, \mathbf{X\gamma}) = \frac{\exp(\mathbf{X\gamma})}{1 + \exp(\mathbf{X\beta}) + \exp(\mathbf{X\gamma})} \tag{11.13b}$$

and the probability $P(Y = 0 \,|\, \mathbf{X\alpha})$ is:

$$P(Y = 0) = \frac{1}{1 + \exp(\mathbf{X\beta}) + \exp(\mathbf{X\gamma})} \tag{11.13c}$$

The pair-wise likelihood functions of the above multinomial logit model against the reference category are:

$$L(\mathbf{\beta}) = \prod_{i=1}^{N} \left[\frac{\exp(\mathbf{X}_i\mathbf{\beta})}{1 + \exp(\mathbf{X}_i\mathbf{\beta})}\right]^{Y_i} \left[1 - \frac{\exp(\mathbf{X}_i\mathbf{\beta})}{1 + \exp(\mathbf{X}_i\mathbf{\beta})}\right]^{(1-Y_i)}, \quad Y_i = 0, 1 \tag{11.14a}$$

$$L(\mathbf{\beta}) = \prod_{i=1}^{N} \left[\frac{\exp(\mathbf{X}_i\mathbf{\gamma})}{1 + \exp(\mathbf{X}_i\mathbf{\gamma})}\right]^{Y_i/2} \left[1 - \frac{\exp(\mathbf{X}_i\mathbf{\gamma})}{1 + \exp(\mathbf{X}_i\mathbf{\gamma})}\right]^{[1-(Y_i/2)]}, \quad Y_i = 0, 2 \tag{11.14b}$$

and the log likelihood functions of the above multinomial logit model are:

$$LL(\mathbf{\beta}) = \sum_{i=1}^{N} Y_i Ln\left[\frac{\exp(\mathbf{X}_i\mathbf{\beta})}{1 + \exp(\mathbf{X}_i\mathbf{\beta})}\right] + (1 - Y_i)Ln\left[1 - \frac{\exp(\mathbf{X}_i\mathbf{\beta})}{1 + \exp(\mathbf{X}_i\mathbf{\beta})}\right] \quad Y_i = 0, 1 \tag{11.15a}$$

$$LL(\mathbf{\beta}) = \sum_{i=1}^{N} \frac{Y_i}{2} Ln\left[\frac{\exp(\mathbf{X}_i\mathbf{\gamma})}{1 + \exp(\mathbf{X}_i\mathbf{\gamma})}\right] + \left(1 - \frac{Y_i}{2}\right)Ln\left[1 - \frac{\exp(\mathbf{X}_i\mathbf{\gamma})}{1 + \exp(\mathbf{X}_i\mathbf{\gamma})}\right] \quad Y_i = 0, 2 \tag{11.15b}$$

In general, a multinomial model with $m + 1$ discrete outcomes has m pairs of log 'odds' against the reference category. Let $P_0 = P(Y = 0 \,|\, \mathbf{X\beta}_0)$ be the base

category or reference category, $P_j = P\left(Y = j \mid \mathbf{X}\boldsymbol{\beta}_j^0\right)$ $(j = 1, \ldots m)$ and $\boldsymbol{\beta}_j = \boldsymbol{\beta}_j^0 - \boldsymbol{\beta}_0$ $(j = 1, \ldots m)$. Then these m pairs of log odds are:

$$\mathrm{Ln}\left(\frac{P_1}{P_0}\right) = \mathbf{X}\boldsymbol{\beta}_1 \tag{11.16a}$$

$$\mathrm{Ln}\left(\frac{P_2}{P_0}\right) = \mathbf{X}\boldsymbol{\beta}_2 \tag{11.16b}$$

$$\ldots$$

$$\mathrm{Ln}\left(\frac{P_m}{P_0}\right) = \mathbf{X}\boldsymbol{\beta}_m \tag{11.16c}$$

The probability $P(Y = 1 \mid \mathbf{X}\boldsymbol{\beta}_1), P(Y = 2 \mid \mathbf{X}\boldsymbol{\beta}_2), \ldots P(Y = m \mid \mathbf{X}\boldsymbol{\beta}_m)$ can be jointly solved as:

$$P(Y = 1 \mid \mathbf{X}\boldsymbol{\beta}_1) = \frac{\exp(\mathbf{X}\boldsymbol{\beta}_1)}{1 + \sum_{j=1}^{m} \exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)} \tag{11.17a}$$

$$P(Y = 2 \mid \mathbf{X}\boldsymbol{\beta}_2) = \frac{\exp(\mathbf{X}\boldsymbol{\beta}_2)}{1 + \sum_{j=1}^{m} \exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)} \tag{11.17b}$$

$$\ldots$$

$$P(Y = m \mid \mathbf{X}\boldsymbol{\beta}_m) = \frac{\exp(\mathbf{X}\boldsymbol{\beta}_m)}{1 + \sum_{j=1}^{m} \exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)} \tag{11.17c}$$

$$P(Y = 0) = \frac{1}{1 + \sum_{j=1}^{m} \exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)} \tag{11.17d}$$

The pair-wise likelihood functions of the above general multinomial logit model against the reference category are:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_1)}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta}_1)}\right]^{Y_i} \left[1 - \frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_1)}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta}_1)}\right]^{(1-Y_i)}, \quad Y_i = 0, 1 \tag{11.18a}$$

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_2)}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta}_2)}\right]^{Y_i/2} \left[1 - \frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_2)}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta}_2)}\right]^{[1-(Y_i/2)]}, \quad Y_i = 0, 2 \tag{11.18b}$$

$$\ldots$$

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_m)}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta}_m)}\right]^{Y_i/m} \left[1 - \frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_m)}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta}_m)}\right]^{[1-(Y_i/m)]}, \quad Y_i = 0, m \tag{11.18c}$$

and the log likelihood functions of the above general multinomial logit model are:

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{N} Y_i \text{Ln}\left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_1)}{1+\exp(\mathbf{X}_i\boldsymbol{\beta}_1)}\right] + (1-Y_i)\text{Ln}\left[1 - \frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_1)}{1+\exp(\mathbf{X}_i\boldsymbol{\beta}_1)}\right]$$

$$Y_i = 0,1 \tag{11.19a}$$

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{Y_i}{2}\text{Ln}\left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_2)}{1+\exp(\mathbf{X}_i\boldsymbol{\beta}_2)}\right] + \left(1-\frac{Y_i}{2}\right)\text{Ln}\left[1 - \frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_2)}{1+\exp(\mathbf{X}_i\boldsymbol{\beta}_2)}\right]$$

$$Y_i = 0,2 \tag{11.19b}$$

$$\ldots$$

$$LL(\boldsymbol{\beta}) = \sum_{i=1}^{N} \frac{Y_i}{m}\text{Ln}\left[\frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_m)}{1+\exp(\mathbf{X}_i\boldsymbol{\beta}_m)}\right] + \left(1-\frac{Y_i}{m}\right)\text{Ln}\left[1 - \frac{\exp(\mathbf{X}_i\boldsymbol{\beta}_m)}{1+\exp(\mathbf{X}_i\boldsymbol{\beta}_m)}\right]$$

$$Y_i = 0,m \tag{11.19c}$$

11.3. Ordered probit and logit

Discrete variables that can be ordered are ordinal variables. Ordered probit and ordered logit are two models used to analyse ordinal dependent variables. Typical ordinal cases are formatted responses to questionnaire surveys, such as the following:

Strongly Disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
1	2	3	4	5

The scaling need not be 1 to 5 and the increment need not be one. It can be, for example, $-2, -1, 0, 1, 2$; or 0, 5, 10.

Let Y^* be an unobserved latent variable that underlies the observed ordinal variable Y in the following way:

$$Y^* = \mathbf{X}\boldsymbol{\beta} + \varepsilon \tag{11.20a}$$

$$Y = i, \quad \text{if } \omega_{i-1} \leq Y^* < \omega_i, \quad i = 1,\ldots k \tag{11.20b}$$

where ω_i is the cut-point. With an ordered probit, the probability of the ith choice or category is chosen is:

$$P(Y = i \mid \mathbf{X}\boldsymbol{\beta}) = P(\mathbf{X}\boldsymbol{\beta} + \varepsilon < \omega_i) - P(\mathbf{X}\boldsymbol{\beta} + \varepsilon < \omega_{i-1})$$

$$= P(\varepsilon < -\mathbf{X}\boldsymbol{\beta} + \omega_i) - P(\varepsilon < -\mathbf{X}\boldsymbol{\beta} + \omega_{i-1})$$

$$= \Phi(-\mathbf{X}\boldsymbol{\beta} + \omega_i) - \Phi(-\mathbf{X}\boldsymbol{\beta} + \omega_{i-1}) \tag{11.21}$$

where $\Phi(z)$ denotes the cumulative normal distribution. The cumulative probability for $Y \leq i$, conditional on $\mathbf{X\beta}$, is:

$$P(Y \leq i | \mathbf{X\beta}) = \sum_{j=1}^{i} P(Y = j | \mathbf{X\beta})$$

$$= \Phi(-\mathbf{X\beta} + \omega_1) + \Phi(-\mathbf{X\beta} + \omega_2) - \Phi(-\mathbf{X\beta} + \omega_1)$$

$$+ \cdots + \Phi(-\mathbf{X\beta} + \omega_i) - \Phi(-\mathbf{X\beta} + \omega_{i-1}) = \Phi(-\mathbf{X\beta} + \omega_i) \qquad (11.22)$$

While the cumulative probability for $Y > i$, conditional on $\mathbf{X\beta}$, is:

$$P(Y > i | \mathbf{X\beta}) = 1 - P(Y \leq i | \mathbf{X\beta}) = \sum_{j=i+1}^{k} P(Y = j | \mathbf{X\beta}) = \Phi(\mathbf{X\beta} - \omega_i)$$

$$(11.23)$$

Using ordered logit for odinal variables, we refer to odds as expected. The odds are the logit of the cumulative probability for $Y \leq i$ or the ratio of the cumulative probability for $Y \leq i$ to the cumulative probability for the rest:

$$\text{Logit}[P(Y \leq i | \mathbf{X\beta})] = \frac{P(Y \leq i | \mathbf{X\beta})}{P(Y > i | \mathbf{X\beta})} = \frac{\sum_{j=1}^{i} P(Y = j | \mathbf{X\beta})}{\sum_{j=i+1}^{k} P(Y = j | \mathbf{X\beta})} \qquad (11.24)$$

and the log odds are:

$$\text{Ln}\{\text{Logit}[P(Y \leq i | \mathbf{X\beta})]\} = \text{Ln}\left[\frac{P(Y \leq i | \mathbf{X\beta})}{P(Y > i | \mathbf{X\beta})}\right]$$

$$= \text{Ln}\left[\sum_{j=1}^{i} P(Y = j | \mathbf{X\beta})\right] - \text{Ln}\left[\sum_{j=i+1}^{k} P(Y = j | \mathbf{X\beta})\right] = \omega_i - \mathbf{X\beta}$$

$$(11.25)$$

We also write the odds and log odds in the following way for the purpose to contrast the multinomial logit model and ordered logit model:

$$\text{Ln}\{\text{Logit}[P(Y > i | \mathbf{X\beta})]\} = \text{Ln}\left[\frac{P(Y > i | \mathbf{X\beta})}{P(Y \leq i | \mathbf{X\beta})}\right]$$

$$= \text{Ln}\left[\sum_{j=i+1}^{k} P(Y = j | \mathbf{X\beta})\right] - \text{Ln}\left[\sum_{j=1}^{i} P(Y = j | \mathbf{X\beta})\right] = \mathbf{X\beta} - \omega_i$$

$$(11.26)$$

Note that the allocation of $0, 1, 2, \ldots$ is arbitrary in multinomial logit models, i.e. we can code 0 for red, 1 for blue, and 2 for yellow; but we can also code 0 for

blue, 1 for yellow, and 2 for red. However, we cannot alter $0, 1, 2, \ldots$ arbitrarily. e.g. for four colours of sky blue, light blue, blue and dark blue, we may code them this way: 0 for sky blue, 1 for light blue, 2 for blue and 3 for dark blue. We may also code them that way: 0 for dark blue, 1 for blue, 2 for light blue and 3 for sky blue. But we cannot code them like this: 0 for blue, 1 for sky blue, 2 for light blue and 3 dark blue. Another example of contrast is coffee, tea and chocolate on the one hand, and a short latte, a tall latte and a grande latte on the other hand. For the former it can be a case of a multinomial logit model and, for the latter, a case of an ordered logit model.

11.4. Marginal effects

There is no need to examine marginal effects for linear probability models that we have experienced, such as linear regression. This is because the marginal effect in such models is constant. The marginal effect measures the effect of a one unit change in one of the explanatory variables on the dependent variable, holding all other explanatory variables constant. In linear probability models such as linear regression, this is simply the coefficient of the explanatory variable. For binary choice models, the probability of the dependent variable being one is a non-linear function of its explanatory variables, though the underlying process of the observed binary variable, supposed to be an unobserved latent variable, is a linear function of explanatory variables. This applies to discrete choice models in general, including multinomial and ordered choice models. Consequently, the marginal effect in these models is non-constant. The marginal effect changes and is specific to the given value of the explanatory variable.

Recalling the probit model of equation (11.2), the marginal effect of the probit model can be derived as follows:

$$\frac{\partial P(Y=1\,|\,\mathbf{X}\boldsymbol{\beta})}{\partial \mathbf{X}} = \frac{\partial\left[\Phi(\mathbf{X}\boldsymbol{\beta})\right]}{\partial \mathbf{X}} = \phi(\mathbf{X}\boldsymbol{\beta})\boldsymbol{\beta}' \tag{11.27}$$

For example, if $\mathbf{X} = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ is a 1×2 row vector of explanatory variables, $\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ is a 2×1 column vector of coefficients, then:

$$\frac{\partial P(Y=1\,|\,\mathbf{X}\boldsymbol{\beta})}{\partial \mathbf{X}}$$

$$= \begin{bmatrix} \dfrac{\partial P(Y=1\,|\,\beta_1 X_1 + \beta_2 X_2)}{\partial X_1} & \dfrac{\partial P(Y=1\,|\,\beta_1 X_1 + \beta_2 X_2)}{\partial X_2} \end{bmatrix}$$

$$= \begin{bmatrix} \dfrac{\partial\left[\Phi(\beta_1 X_1 + \beta_2 X_2)\right]}{\partial X_1} & \dfrac{\partial\left[\Phi(\beta_1 X_1 + \beta_2 X_2)\right]}{\partial X_2} \end{bmatrix}$$

$$= \phi(\beta_1 X_1 + \beta_2 X_2)\begin{bmatrix} \beta_1 & \beta_2 \end{bmatrix} \tag{11.28}$$

It is the derivative of the probability of $Y = 1$, conditioning on the explanatory variables, with respect to each of the explanatory variables. Since $\phi(\mathbf{X}\boldsymbol{\beta})$ generally

changes when the explanatory variables change, marginal effects or marginal effect coefficients, measured by $\phi(\mathbf{X}\boldsymbol{\beta}) \times \boldsymbol{\beta}$, are generally non-constant. Since $\phi(\mathbf{X}\boldsymbol{\beta})$ is usually smaller than one, marginal effects or marginal effect coefficients are usually smaller that the estimated coefficients of $\boldsymbol{\beta}$. Therefore, the effect of the explanatory variable on the dependent variable would be overvalued without the adjustment for marginal effects.

Referring to equation (11.5), the marginal effect of the logit model is derived as:

$$
\begin{aligned}
\frac{\partial P(Y=1\,|\,\mathbf{X}\boldsymbol{\beta})}{\partial \mathbf{X}} &= \frac{\partial \left[\dfrac{\exp(\mathbf{X}\boldsymbol{\beta})}{1+\exp(\mathbf{X}\boldsymbol{\beta})} \right]}{\partial \mathbf{X}} \\[2mm]
&= \frac{\exp(\mathbf{X}\boldsymbol{\beta})[1+\exp(\mathbf{X}\boldsymbol{\beta})]\boldsymbol{\beta}' - \exp(\mathbf{X}\boldsymbol{\beta})\exp(\mathbf{X}\boldsymbol{\beta})\boldsymbol{\beta}'}{[1+\exp(\mathbf{X}\boldsymbol{\beta})]^2} \\[2mm]
&= \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{[1+\exp(\mathbf{X}\boldsymbol{\beta})]^2}\boldsymbol{\beta}' \qquad\qquad (11.29)
\end{aligned}
$$

Similar to the probit case, marginal effects or marginal effect coefficients are smaller than the estimated coefficients of $\boldsymbol{\beta}$ due to the fact that $\exp(\mathbf{X}\boldsymbol{\beta})/[1+\exp(\mathbf{X}\boldsymbol{\beta})]^2$ is smaller than one, and the effect of the explanatory variable on the dependent variable would be overvalued without the adjustment for marginal effects.

From the above analysis, it is straightforward to derive marginal effects or marginal effect coefficients for the multinomial logit model:

$$
\begin{aligned}
\frac{\partial P(Y=1\,|\,\mathbf{X}\boldsymbol{\beta}_1)}{\partial \mathbf{X}} &= \frac{\partial \left[\dfrac{\exp(\mathbf{X}\boldsymbol{\beta}_1)}{1+\sum_{j=1}^{m}\exp(\mathbf{X}\boldsymbol{\beta}_j)} \right]}{\partial \mathbf{X}} \\[3mm]
&= \frac{\exp(\mathbf{X}\boldsymbol{\beta}_1)\left[1+\sum_{j=1}^{m}\exp(\mathbf{X}\boldsymbol{\beta}_j)\right]\boldsymbol{\beta}_1' - \exp(\mathbf{X}\boldsymbol{\beta}_1)\left[\sum_{j=1}^{m}\exp(\mathbf{X}\boldsymbol{\beta}_j)\boldsymbol{\beta}_j'\right]}{\left[1+\sum_{j=1}^{m}\exp(\mathbf{X}\boldsymbol{\beta}_j)\right]^2} \\[3mm]
&= \frac{\exp(\mathbf{X}\boldsymbol{\beta}_1)\,\boldsymbol{\beta}_1' + \exp(\mathbf{X}\boldsymbol{\beta}_1)\left[\sum_{j=1}^{m}\exp(\mathbf{X}\boldsymbol{\beta}_j)\,(\boldsymbol{\beta}_1' - \boldsymbol{\beta}_j')\right]}{\left[1+\sum_{j=1}^{m}\exp(\mathbf{X}\boldsymbol{\beta}_j)\right]^2} \qquad (11.30\mathrm{a})
\end{aligned}
$$

$$
\frac{\partial P(Y=2\,|\,\mathbf{X}\boldsymbol{\beta}_2)}{\partial \mathbf{X}} = \frac{\exp(\mathbf{X}\boldsymbol{\beta}_2)\boldsymbol{\beta}_2' + \exp(\mathbf{X}\boldsymbol{\beta}_2)\left[\sum_{j=1}^{m}\exp(\mathbf{X}\boldsymbol{\beta}_j)\,(\boldsymbol{\beta}_2' - \boldsymbol{\beta}_j')\right]}{\left[1+\sum_{j=1}^{m}\exp(\mathbf{X}\boldsymbol{\beta}_j)\right]^2}
$$

$$
(11.30\mathrm{b})
$$

$\cdots$

$$\frac{\partial P(Y=m\,|\,\mathbf{X}\boldsymbol{\beta}_m)}{\partial \mathbf{X}} = \frac{\exp(\mathbf{X}\boldsymbol{\beta}_2)\boldsymbol{\beta}'_m + \exp(\mathbf{X}\boldsymbol{\beta}_m)\left[\sum_{j=1}^{m}\exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)\left(\boldsymbol{\beta}'_m - \boldsymbol{\beta}'_j\right)\right]}{\left[1 + \sum_{j=1}^{m}\exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)\right]^2}$$

(11.30c)

$$\frac{\partial P(Y=0)}{\partial \mathbf{X}} = \frac{-\sum_{j=1}^{m}\exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)\boldsymbol{\beta}'_j}{\left[1 + \sum_{j=1}^{m}\exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)\right]^2}$$

(11.30d)

These marginal effects or marginal effect coefficients for the multinomial logit model can also be expressed in a slightly different way as follows:

$$\frac{\partial P(Y=1\,|\,\mathbf{X}\boldsymbol{\beta}_1)}{\partial \mathbf{X}} = \frac{\exp(\mathbf{X}\boldsymbol{\beta}_1)\boldsymbol{\beta}'_1 + \exp(\mathbf{X}\boldsymbol{\beta}_1)\left[\sum_{j=1}^{m}\exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)\left(\boldsymbol{\beta}'_1 - \boldsymbol{\beta}'_j\right)\right]}{\left[1 + \sum_{j=1}^{m}\exp\left(\mathbf{X}\boldsymbol{\beta}_j\right)\right]^2}$$

$$= P(Y=1\,|\,\mathbf{X}\boldsymbol{\beta}_1)\boldsymbol{\beta}'_1 - P(Y=1\,|\,\mathbf{X}\boldsymbol{\beta}_1)\sum_{j=1}^{m}P\left(Y=j\,|\,\mathbf{X}\boldsymbol{\beta}_j\right)\boldsymbol{\beta}'_j$$

$$= P(Y=1\,|\,\mathbf{X}\boldsymbol{\beta}_1)\left(\boldsymbol{\beta}'_1 - \boldsymbol{\beta}'_j\right)$$

(11.31a)

$$\frac{\partial P(Y=2\,|\,\mathbf{X}\boldsymbol{\beta}_2)}{\partial \mathbf{X}} = P(Y=2\,|\,\mathbf{X}\boldsymbol{\beta}_2)\left(\boldsymbol{\beta}'_2 - \boldsymbol{\beta}'_j\right)$$

(11.31b)

$$\cdots$$

$$\frac{\partial P(Y=2\,|\,\mathbf{X}\boldsymbol{\beta}_m)}{\partial \mathbf{X}} = P(Y=2\,|\,\mathbf{X}\boldsymbol{\beta}_m)\left(\boldsymbol{\beta}'_m - \boldsymbol{\beta}'_j\right)$$

(11.31c)

The marginal effects of the ordered probit can be derived as follows:

$$\frac{\partial\left[P(Y\leq i\,|\,\mathbf{X}\boldsymbol{\beta})\right]}{\partial \mathbf{X}} = \frac{\partial\left[\Phi(-\mathbf{X}\boldsymbol{\beta} + \omega_i)\right]}{\partial \mathbf{X}} = \phi(-\mathbf{X}\boldsymbol{\beta} + \omega_i)\boldsymbol{\beta}'$$

(11.32a)

$$\frac{\partial\left[P(Y>i\,|\,\mathbf{X}\boldsymbol{\beta})\right]}{\partial \mathbf{X}} = \frac{\partial\left[\Phi(\mathbf{X}\boldsymbol{\beta} - \omega_i)\right]}{\partial \mathbf{X}} = \phi(\mathbf{X}\boldsymbol{\beta} - \omega_i)\boldsymbol{\beta}'$$

(11.32b)

While the marginal effects of the ordered logit are:

$$\frac{\partial\left[P(Y\leq i\,|\,\mathbf{X}\boldsymbol{\beta})\right]}{\partial \mathbf{X}} = \frac{\partial\left[\dfrac{\exp(\omega_i - \mathbf{X}\boldsymbol{\beta})}{1 + \exp(\omega_i - \mathbf{X}\boldsymbol{\beta})}\right]}{\partial \mathbf{X}} = \frac{\exp(\omega_i - \mathbf{X}\boldsymbol{\beta})}{\left[1 + \exp(\omega_i - \mathbf{X}\boldsymbol{\beta})\right]^2}\boldsymbol{\beta}'$$

(11.33a)

$$\frac{\partial\left[P(Y>i\,|\,\mathbf{X}\boldsymbol{\beta})\right]}{\partial \mathbf{X}} = \frac{\partial\left[\dfrac{\exp(\mathbf{X}\boldsymbol{\beta} - \omega_i)}{1 + \exp(\mathbf{X}\boldsymbol{\beta} - \omega_i)}\right]}{\partial \mathbf{X}} = \frac{\exp(\mathbf{X}\boldsymbol{\beta} - \omega_i)}{\left[1 + \exp(\mathbf{X}\boldsymbol{\beta} - \omega_i)\right]^2}\boldsymbol{\beta}'$$

(11.33b)

11.5. Examples and cases

Example 11.1

The growing use of the Internet for trading and other commercial purposes, within a fairly short spell of the existence of the Internet, is phenomenal. What prompts customers to buy on-line has become a relevant as well as an interesting topic of research on e-commerce recently, though people had rarely taken this on-line shopping issue seriously, if on-line shopping had been considered an issue at all a few years earlier. In a study by Koyuncu and Bhattacharya (2004), the effects of selected factors on on-line shopping behaviour have been examined, using binomial and multinomial logistic models. There are a number of risks and benefits associated with e-commerce, on the demand side as well as on the supply side. Quickness of transactions and low prices are identified as benefits, and payment risk and delivery issues are identified as risk factors in the paper. The authors hypothesise that quickness of transactions and lower prices contribute to an individual's decision to purchase on the Internet positively while payment risk and delivery issues contribute to an individual's decision to purchase on the Internet negatively. Their results from estimating both binomial and multinomial logistic models indicate that the hypothesis can be accepted. Further, they use the multinomial logistic model to estimate the frequency of on-line shopping by individuals, and claim to have found that the frequency of on-line shopping increases with the benefits of e-commerce and decreases with its risks, a seemingly obvious result.

   A survey data set collected by Georgia Institute of Technology in April 1998 is used in the study. The final sample consists of 1842 individual Internet users in the US after eliminating observations with missing data. The data set contains information on each individual's demographic and economic characteristics as well as the individual's degrees of agreement on particular on-line issues of price, quickness, payment risk, and delivery. The demographic and economic characteristics used in the study include age, education, gender, income and experience with the Internet. Among these 1842 Internet users, 1504 are on-line buyers and 338 are non-buyers. Male on-line shoppers account for 54 per cent for the whole sample, 58 per cent for the buyer group, and 39 per cent for the non-buyer group respectively. The coding for the survey questions is as follows. Age is in 11 bands ranging from 1 for under 21 to 11 for over 65; Income is also in 11 bands, from 1 for under $10,000 to 11 for over $100,000 annual household income; Education is in eight bands; With regard to gender, 1 is for male and 0 for female and all other responses; Internet experience is in nine bands, 1 for under 6 months and 9 for over 7 years; The four questions regarding on-line shopping issues of price, quickness, payment risk, and delivery are on a 1–5 scale, with

1 being strongly disagree, 2 being disagree, 3 being neither disagree no agree, 4 being agree and 5 being strongly agree.

First, they use a binomial logistic model to evaluate the impact of an individual's reaction to price, quickness, payment risk, and delivery issues on that individual's on-line shopping decision. Table 11.1 reports the results from estimating the binomial model. The second and third columns are for their benchmark model, which includes only the variables representing demographic and economic factors and on-line shopping experience. The last two columns are for their full model, further including individuals' degrees of agreement on the four survey questions regarding price, quickness, payment risk, and delivery. Both the coefficients and marginal effects coefficients are reported. All the coefficients, except that for the age variable, are statistically significant at the 10 per cent level or higher for both models. Nevertheless, the level of significance of the coefficients for the demographic and economic character variables falls when the on-line shopping character variables are included. According to these results,

Table 11.1 Binomial logistic estimation of on-line shopping

	Coefficient	*Marginal effect*	*Coefficient*	*Marginal effect*
Constant	−1.233***	−0.161***	0.110	0.011
	(0.268)	(0.035)	(0.577)	(0.058)
Age	0.042	0.005	0.024	0.002
	(0.026)	(0.003)	(0.028)	(0.003)
Income	0.075***	0.097***	0.052**	0.005**
	(0.023)	(0.003)	(0.025)	(0.002)
Education	0.140***	0.018***	0.114**	0.012**
	(0.050)	(0.007)	(0.054)	(0.006)
Gender	0.527***	0.069***	0.423***	0.043***
	(0.129)	(0.017)	(0.144)	(0.015)
Experience	0.266***	0.035***	0.204***	0.021***
	(0.033)	(0.004)	(0.035)	(0.003)
Quickness			0.394***	0.040***
			(0.069)	(0.007)
Price			0.297***	0.030***
			(0.076)	(0.008)
Risk			−0.687***	−0.069***
			(0.074)	(0.007)
Delivery			−0.186**	−0.019**
			(0.087)	(0.009)

Standard errors in parentheses.
*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.
Marginal effects evaluated at the sample means.

income, education and internet experience all have a positive effect on the on-line shopping decision. That is, the higher the household income, the more likely the individual would shop on-line; the higher the individual's education, the more likely that individual would shop on-line; the more internet experience the individual has, the more likely that individual would shop on-line. However, it is unclear why being a male is more likely to shop on-line. The effect of the four on-line shopping issues is as expected. The advantages of better prices and time-saving offered by on-line shopping induce individuals to buy on-line, with the two coefficients for the variables representing these two aspects of on-line shopping being positive and highly significant at the 1 per cent level. The risk of making on-line payment is confirmed a concern for not using the Internet for shopping, with a highly significant negative coefficient for the variable. To a lesser extent, a long period for delivering the item purchased via the Internet has a negative effect on the on-line shopping decision, with the coefficient for the variable being negatively significant at the 5 per cent level.

They then examine the effect of the four on-line issues on individuals' on-line purchasing frequencies. They set five different categories according to the frequency with which on-line shopping takes place. 0 denotes the category for individuals who do not make on-line shopping at all; 1 denotes the category for individuals who do on-line shopping less than once each month; 2 denotes the category for individuals who do on-line shopping about once each month; 3 denotes the category for individuals who do on-line shopping several times each month; and 4 denotes the category for individuals who do on-line shopping about once per week or more frequently. In their multinomial logistic analysis, the base category is those individuals who do not shop on-line at all. Table 11.2 reports the results from estimating the multinomial model.

Unlike binomial models where the sign of the marginal effects coefficient is always the same as that of the coefficient, the sign of the marginal effects coefficient can be different from that of its corresponding coefficient. There are several examples in Table 11.2. Equations (11.30) and (11.31) reveal the fact and reason why the sign of the marginal effects coefficient may be different from that of the coefficient for multinomial models. Let us pay attention to panel B, the results of the full model. The marginal effects coefficient has changed the sign *vis-à-vis* its corresponding coefficient for all the four on-line shopping variables, as well as income, education and gender, for $\text{Prob}(y = 1)$, the category for individuals who do on-line shopping less than once each month. The sign of the marginal effects coefficient is not desirable, while the sign of the coefficient is as expected, for all the four on-line shopping variables of price, quickness, payment risk and delivery for this category. In their study, the advantages of better prices

Table 11.2 Multinomial logistic estimation of on-line shopping

	Prob(y = 1)		Prob(y = 2)		Prob(y = 3)		Prob(y = 4)	
	Coef	Mgnl effect	Coef	Mgnl effect	Coef	Mgnl effect	Coef	Mgnl effect
A								
Constant	−1.234***	0.351***	−3.059***	−0.253***	−4.380***	−0.285***	−6.995***	−0.127***
	(0.284)	(0.051)	(0.340)	(0.054)	(0.415)	(0.051)	(0.743)	(0.035)
Age	0.051*	0.008*	0.019	−0.003	0.026	−0.001	0.096	0.002
	(0.027)	(0.005)	(0.032)	(0.004)	(0.038)	(0.003)	(0.062)	(0.001)
Income	0.045*	−0.009**	0.097***	0.008**	0.133***	0.008***	0.186***	0.003**
	(0.024)	(0.004)	(0.027)	(0.004)	(0.031)	(0.003)	(0.050)	(0.001)
Education	0.089*	−0.016*	0.190***	0.016*	0.264***	0.017***	0.231*	0.003
	(0.053)	(0.009)	(0.061)	(0.008)	(0.072)	(0.006)	(0.119)	(0.003)
Gender	0.415***	−0.017	0.611***	0.038*	0.924***	0.056***	0.242	−0.005
	(0.137)	(0.024)	(0.156)	(0.021)	(0.187)	(0.018)	(0.297)	(0.007)
Experience	0.227***	0.001	0.315***	0.021***	0.287***	0.007**	0.471***	0.006***
	(0.035)	(0.005)	(0.039)	(0.005)	(0.044)	(0.004)	(0.069)	(0.002)
B								
Constant	−0.197	0.443***	−1.675**	−0.152	−4.174***	−0.306***	−7.411***	−0.110***
	(0.590)	(0.103)	(0.707)	(0.093)	(0.855)	(0.075)	(1.424)	(0.040)
Age	0.031	0.006	0.001	−0.005	0.021	0.000	0.096	0.001
	(0.029)	(0.005)	(0.035)	(0.005)	(0.042)	(0.003)	(0.065)	(0.001)
Income	0.033	−0.011**	0.086***	0.008**	0.128***	0.007***	0.179***	0.002**
	(0.025)	(0.004)	(0.030)	(0.004)	(0.035)	(0.002)	(0.053)	(0.001)
Education	0.078	−0.021**	0.193***	0.018**	0.275***	0.015***	0.238*	0.002
	(0.056)	(0.010)	(0.067)	(0.009)	(0.079)	(0.006)	(0.125)	(0.002)

Continued

Table 11.2 continued

	Prob(y = 1)		Prob(y = 2)		Prob(y = 3)		Prob(y = 4)	
	Coef	Mgnl effect	Coef	Mgnl effect	Coef	Mgnl effect	Coef	Mgnl effect
Gender	0.361**	−0.018	0.505***	0.026	0.828***	0.042***	0.210	−0.003
	(0.147)	(0.026)	(0.176)	(0.023)	(0.209)	(0.016)	(0.317)	(0.005)
Experience	0.188***	0.003	0.238***	0.014***	0.195***	0.001	0.379***	0.003***
	(0.036)	(0.006)	(0.042)	(0.005)	(0.049)	(0.003)	(0.074)	(0.001)
Quickness	0.314***	−0.037***	0.514***	0.031**	0.809***	0.042***	0.984***	0.010***
	(0.070)	(0.014)	(0.089)	(0.012)	(0.112)	(0.010)	(0.197)	(0.004)
Price	0.218***	−0.043***	0.493***	0.046***	0.574***	0.027***	0.633***	0.006**
	(0.078)	(0.014)	(0.097)	(0.013)	(0.116)	(0.009)	(0.189)	(0.003)
Risk	−0.522***	0.065***	−1.008***	−0.087***	−1.129***	−0.047***	−1.158***	−0.009**
	(0.077)	(0.013)	(0.090)	(0.013)	(0.104)	(0.009)	(0.159)	(0.004)
Delivery	−0.138	0.023	−0.303***	−0.029**	−0.303**	−0.012	−0.417**	−0.004
	(0.089)	(0.015)	(0.104)	(0.013)	(0.119)	(0.008)	(0.182)	(0.003)

Standard errors in parentheses.
* significant at the 10 per cent level; ** significant at the 5 per cent level; *** significant at the 1 per cent level.
Marginal effects evaluated at the sample means.

and quickness offered by on-line shopping still attract the individuals in this category to shop on-line as suggested by the positive coefficients for these two variables, significant at the 1 per cent level. But these advantages attract individuals to shop on-line more frequently, resulting in negative marginal effects of these two variables. This is in relation to the more frequent on-line buyers in categories 2, 3 and 4, not in relation to the non-buyers in the base category 0, and should not be misinterpreted. Koyuncu and Bhattacharya (2004), by basing their explanations on the marginal effects coefficients exclusively, have encountered apparent trouble. They state with puzzlement: 'Unexpectedly, we found that the quickness and price variables make negative while the risk variable makes positive and statistically significant contributions to the probability of purchasing from the Internet for those individuals who do on-line shopping less than once each month', which has misinterpreted their own empirical results. For the remaining three categories, the sign of the marginal effects coefficient of all the four variables of price, quickness, payment risk and delivery is as expected. An ordered logit would be more appropriate to model this on-line shopping case where the dependent variable is the frequency with which individuals shop on-line. If they had chosen an ordered logit model, the marginal effects coefficients for the price and quickness variables for category 1 would have been positive. There would have been no misinterpretation and confusion then.

Example 11.2

Mergers and acquisitions (M&As), or corporate takeovers, are among the most important investment decisions made by acquiring firms and, consequently, have attracted immense attention from academia and professionals alike. Over years, many studies have been carried out to examine the motivations for M&As and the factors that have an influence on a takeover decision made by acquiring firms. In an empirical study to predict which firms become takeover targets, Espahbodi and Espahbodi (2003) employ four models of binary choice of logit, probit, discriminant, and recursive partitioning to examine a dataset gathered from *Mergers and Acquisitions*, a bi-monthly periodical. This section focuses on their logit and probit results while briefly remarking on recursive partitioning. They have examined the November–December, September–October, and July–August issues in 1997 only, claiming that it is to minimise the time series distortion

Continued

in their models. Regardless, there are a few approaches available to deal with the effect on model estimation of the changing economic environment and the resulting operating characteristics of firms; or indeed, the issues associated with these changes should be addressed not avoided. Utilities and financial institutions are dropped from the study. The final sample consists of 133 target firms, deemed by them to be large enough for analysis, probably large enough for statistical considerations.

After lengthy discussions of 18 financial variables and additionally several non-financial variables that may identify a takeover target firm, Espahbodi and Espahbodi (2003) adopt four of them, namely, free cash flow over total assets, existence of golden parachutes, the State of Delaware incorporation dummy variable and market value of equity over total firm value, in their logit and probit analysis. The State of Delaware incorporation dummy variable is included since Delaware has tougher takeover laws than other states. As a result, firms incorporated in Delaware may be subject to lower probabilities of a hostile takeover. Whereas, friendly takeovers may increase as such firms can use the anti-takeover laws as a leverage to extract a higher price for their firms. The dummy takes a value of 1 if a firm is incorporated in Delaware and 0 otherwise. The sign of the coefficient for the dummy and its significance are thus uncertain, claimed by the authors.

The results from estimating the logit model and probit model are presented in Table 11.3. Columns 2 and 3 are for the logit model and columns 4 and 5 for the probit model results. Free cash flow over total assets is found to be positively related to the likelihood of a takeover of the firm significantly by both models, which is consistent with the hypotheses with regard to

Table 11.3 Estimation of takeovers by the logit and probit models

	Logit		Probit	
	Coefficient	*t-stat*	*Coefficient*	*t-stat*
Free cash flow over total assets	1.428***	2.76	0.813***	2.86
Existence of golden parachutes	0.612*	1.95	0.366*	1.91
State of Delaware incorporation dummy variable	0.340	1.61	0.204*	1.65
Market value of equity over total firm value	−0.606	−1.36	−0.353	−1.33
Constant	−0.811		0.508	
Likelihood ratio	−285.4		−285.2	
R^2	19.39		19.49	

*significant at the 10 per cent level; ***significant at the 1 per cent level.

Table 11.4 Classifications of target and non-target firms

	Logit			*Probit*			
	Classified as			*Classified as*			
	Target	Non-target	Total	Target	Non-target	Total	
Target	83	50	133	82	51	133	Target
Non-target	153	232	385	147	238	385	Non-target
Total	236	282	518	229	289	518	Total
Correct classification	62.4%	60.3%		61.7%	61.8%		Correct classification

(Left side labelled *Actual*; right side labelled *Actual*)

investment opportunities and agency costs. Existence of golden parachutes is also expected to increase the likelihood of takeover under both the incentive alignment and management entrenchment hypotheses, which is confirmed by the results of both models. A higher ratio of market value of equity over total firm value tends to reduce the likelihood of a takeover of the firm, with a negative but insignificant coefficient in both models, albeit consistent with the growth option hypothesis. The Delaware dummy variable has a positive coefficient that is marginally significant. The result indicates that the target firms are using the tough anti-takeover laws in Delaware to negotiate a higher price for stockholders, claimed by the authors.

The classification of target and non-target firms, or prediction of takeovers, is presented in Table 11.4, with the left panel by the logit model and the right panel by the probit model. The correct classification rate is 62.4 per cent for target firms and 60.3 per cent for non-target firms by the logit model. That is, only 62.4 per cent of the firms predicted by the model to be a takeover target were actually a target; and only 60.3 per cent of the firms predicted by the model to be non-targets were actually non-targets. The probit model produces similar disappointing results. The correct classification rate is 61.7 per cent for target firms and 61.8 per cent for non-target firms. The authors also use a recursive partitioning model to predict takeover targets, which looks like a decision trees model without feedback mechanism. It is alleged that the accuracy of the classification or prediction is increased to 89.5 per cent for target firms and 88.1 per cent for non-target firms. However, it is unclear how the recursive partitioning model results are compared with the logit and probit results, since they employ considerably different variables in the modelling. Nonetheless, it is not the interest of this section to discuss this method at length.

Example 11.3

One of the most serious debated issues in many countries in the world is an aging population. The implications of aging populations are serious, creating a need to study other related important issues, notably, the retirement system. The choice and provision of retirement systems have attracted the interest of numerous parties, including governments, the general public, employees, employers and academics. In this area of research, Kim and DeVaney (2005) investigate whether older workers choose partial or full retirement, or continue to work full-time. They employ multinomial logistic regression to examine data from the first and fifth waves of the Health and Retirement Study (HRS) collected in 1992 and 2000, an ongoing longitudinal dataset that focuses on the retirement and demographic characteristics of older Americans in the 1990s. The dependent variable of the model is retirement status measured by self-reported retirement or working status and change in the number of hours worked during the 8-year period between Wave 1 and Wave 5. The dependent variable takes the value of 0 for full-time work, 1 for partial retirement and 2 for full retirement, with full-time work being the reference category. The independent variables are divided into five sets of household assets and income, pension and health insurance factors, health status, occupational characteristics, and demographic characteristics. The values and coding of the independent variables in Table 11.5 are described as follows. Variables under household assets and income are continuous and measured in US dollars. Variables under pension and health insurance are discrete with 1 for a 'yes' answer and 0 for 'no'. The first two variables under health status are discrete, taking the value of 1 for a 'yes' answer and 0 for 'no'; it is obviously the number of occurrences for the next two variables. Variables under occupational characteristics and demographic characteristics are all discrete, being 1 for a 'yes' answer and 0 for 'no', with the exception of age that is measured in years.

   Table 11.5 reports the results of multinomial logistic regression analysis. Let us look at the results for full retirement first. Household investment assets contribute to a decision to take full retirement positively, while older workers with more household debt are more likely to work full-time instead of taking full retirement. Older workers with a defined benefit plan or both defined benefit and defined contribution plans are more likely to retire fully. Health status is found to be negatively related to full retirement – older workers with excellent or very good health are more likely to continue full-time work. Self-employment is negatively related to the likelihood of full retirement versus full-time work, i.e. self-employed workers are more likely to continue work at old ages. The meaning of work is also a contributing factor to the retirement decision. Older workers who consider the meaning of work

Table 11.5 Results of multinomial logistic regression analysis of retirement status

	Retirement status			
	Partial retirement (N = 461)		Full retirement (N = 990)	
	Coefficient	Std error	Coefficient	Std error
Household assets and income				
Ln (Labor income)	−0.0444	0.0283	−0.0021	0.0200
Ln (Unearned income)	0.0112	0.0184	0.0049	0.0026
Ln (Liquid assets)	0.0005	0.0202	0.0013	0.0165
Ln (Investment assets)	0.0174	0.0220	0.0519***	0.0182
Ln (Real assets)	0.0380	0.0238	0.0140	0.0120
Ln (Debt)	0.0128	0.0142	−0.0230***	0.0092
Pension and health insurance				
DB plan only	0.0224	0.1564	0.7971***	0.1289
DC plan only	−0.2464	0.1682	0.2547	0.2402
Both DB and DC plans	−0.3584	0.4222	0.5964**	0.3016
IRA and Keogh plans	−0.0333	0.2155	−0.0740	0.1759
Employee health insurance	0.1741	0.1258	0.2021*	0.1033
Health status				
Excellent/very good health	0.4276	0.2332	−0.4021***	0.1615
Good health	0.4735	0.3233	−0.2362	0.1613
Number of serious conditions	−0.1534	0.1656	−0.1336	0.1216
Number of chronic conditions	0.1220***	0.0434	0.0421	0.0358
Occupational characteristics				
Self-employment	0.3453**	0.1726	−0.3468**	0.1658
Physically demanding work	0.1299	0.1654	0.0692	0.1509
Mentally challenging work	0.1579	0.1590	0.1540	0.1435
Perceived age discrimination	0.1026	0.1678	0.2000	0.1300
Meaning of work is more important	−0.0635	0.1838	−0.3260***	0.0983
Demographic characteristics				
Age	0.2361***	0.0187	0.3088***	0.0162
Male	0.5659***	0.1125	0.4910***	0.1030
Married with working spouse	−0.1870	0.1720	−0.1970	0.1505
Married with non-working spouse	−0.0934	0.1419	−0.1947	0.1479
African American	0.2744	0.2691	0.1375	0.1878
Hispanic	0.0180	0.2343	0.1073	0.1989
Other	−0.4852	0.4618	−0.0793	0.3488
High school	0.21981	0.1984	0.0138	0.1336
College degree	0.4091**	0.1830	−0.1013	0.1513
More than college	0.1198	0.2329	−0.4729***	0.1282
Constant	−9.7652***	2.1180	−7.9884***	1.8460
Log-likelihood	4780.76			

*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at 1 per cent level.

is more important than money are more likely to continue full-time work *vis-à-vis* full retirement. Obviously, age is confirmed to be positively related to the likelihood of full retirement. According to their results, older male workers are more likely to take full retirement than their female counterparts. Kim and DeVaney (2005) argue that the finding probably reflects the fact that women experience more career interruption for family matters and, as a result, they tend to remain in their full-time jobs. Education is found to be related to retirement decisions. Older workers with more education than a college degree are more likely to choose to continue full-time work.

The boundary between partial retirement and full-time work is apparently less clear-cut than that between full retirement and full-time work. Consequently, fewer coefficients in the partial retirement case are statistically significant. Household assets and income do not contribute to partial retirement decisions at all. The number of chronic health conditions is found to be positively related to partial retirement, suggesting that older workers with chronic conditions have difficulties to work full-time but they are willing to work part-time, or more likely to choose partial retirement. In contrast to a decision to choose between full-time work and full retirement, self-employment is found to be positively related to the likelihood of partial retirement versus full-time work. Kim and DeVaney (2005) claim that, because the self-employed have the flexibility to establish their hours, the choice of partial retirement versus full-time work is rational. The results for age and gender are the same as in the case of full retirement versus full retirement. Older workers with a college degree are found to be more likely to partially retire, which is also in contrast to the case of full retirement versus full-time work to a certain extent.

11.6. Empirical literature

Choice is deeply associated with everyday life, be it social or personal, collective or individual. In financial terms, people make choice aimed at achieving higher utility of their work, consumption, savings, investment, and their combinations. Corporations make choices supposedly aimed at maximising shareholder values. Binary choice and discrete choice models such as probit, binomial and multinomial logistic regression have been traditionally applied to social science research, employment studies and the labour market, health services and insurance. Recently, however, there has been growing interest in corporate decision making involving choice explicitly, as well as in financial market investment choice at the micro level and monetary policy at the macro level.

Examining the factors that have an effect on bank switching in small firms in the UK Howorth *et al.* (2003) employ binomial and multinomial logistic regression to identify the characteristics that discriminate between three groups

of small firms. The three groups of firms are classified as a group of firms considering switching banks, a group of firms that had switched banks in the previous three years and a group of firms that had not switched banks and were not considering doing so. They test the hypothesis that some small firms may be 'informationally captured' in that they are tied into their current bank relationship due to difficulties in conveying accurate information about their performance. Their empirical results are not conclusive with regard to this hypothesis. There is some evidence in support of the hypothesis where rapidly changing information and particularly changing technology are characteristics associated with firms that were considering switching but had not switched. However, there is no significant evidence to indicate that superior performing firms are more likely to be 'informationally captured', as growth and perceived business success are both the factors that influence firms to switched banks. Other factors considered to be relevant to bank-switching decisions in the paper include difficulties obtaining finance, dissatisfaction with the service provided, and the use of alternative sources of finance. Reiljan (2002) studies the determinants of foreign direct investment (FDI) in Estonia. The author firstly adopts principal component analysis to establish five major components of determinants of foreign direct investment, and then runs multinomial logistic regression to investigate the impact of these determinants on different groups of foreign investors. The impact of financial information and voluntary disclosures on contributions to not-for-profit organisations is scrutinised by Parsons (2007). Her study consists of two experiments, with one experiment being a survey with potential donors and the other laboratory based experiment. Potential donors were sent, via a direct mail campaign, fund-raising appeals containing varying amounts of financial and non-financial information in order to determine whether individual donors are more likely to contribute when accounting information or voluntary disclosures are provided. Participants in a laboratory experiment were asked to assess the usefulness of the different versions of the fund-raising appeals. It is claimed that the results of logistic regression provides evidence that some donors who have previously donated use financial accounting information when making a donation decision. The results suggest that non-financial service efforts and voluntary accomplishments disclosures are not clear-cut factors for determining whether and how much to give by donors, though these non-financial factors are considered to be useful by participants in the laboratory experiment for making donation decisions. Factors that restrict the use of credit cards in the Gulf Cooperation Council (GCC) countries are examined by Metwally and Prasad (2004). They adopt the logit model and the probit model to analyse the factors considered to determine the probability of using credit cards more frequently in domestic transactions. The study has supposedly covered the GCC countries of Bahrain, Kuwait, Oman, Qatar, Saudi Arabia and the United Arab Emirates, but it appears to be conducted in Qatar only, where a sample of 385 consumers was surveyed. Respondents who hold credit cards were asked to indicate their degree of agreement on 23 statements relating to their reluctance to use credit cards frequently in domestic transactions on a five-point scale. They use principal component analysis to reduce the 23 explanatory

variables as represented by the 23 statements to five factors, and then use the factor scores of the five extracted factors as explanatory variables in logit and probit regression. Their logit and probit regression results suggest that the probability of using credit cards more frequently in domestic transactions in Qatar would be higher, the closer the credit-card price to cash price and the smaller the debt-service ratio. They also examine the effect of demographic variables on the use of credit cards. Their results indicate that there is a high degree of similarity between Qatar and developed countries regarding the effect of these variables on the use of credit cards.

The world has been in an accelerating process of globalisation over the last three decades. Under such an economic environment, companies' overseas market listings have become a rather familiar phenomenon. This provides companies with more opportunities in terms of choice, but creates a new decision-making area for companies to tackle seriously as well. In a paper entitled 'The choice of trading venue and relative price impact of institutional trading: ADRs versus the underlying securities in their local markets', Chakravarty *et al.* (2004) address two issues of the choice of trading venues and the comparison of trading costs across the venues. They identify institutional trading in the US in non-US stocks, i.e. ADRs, from 35 foreign countries and in their respective home markets, using proprietary institutional trading data. Then they use multinomial logistic regression to examine the factors that influence institutions' decisions to trade a cross-listed stock solely in the ADR market versus solely on its home exchange. They allege that stocks with tentatively higher local volume, with non-overlapping trading hours, and with smaller market capitalisation are more likely to be traded on their home exchanges only while less complex decisions are more likely to be executed as ADRs only, relative to stocks that are traded by institutions in both venues. They also claim that the trading cost of ADRs is often higher than that of the equivalent security at home in terms of overall trading costs. Further, their multivariate analysis on institutional trading costs indicates that the cost difference between trading in the security's home country and its respective ADR is smaller for stocks associated with less complex trades, relatively lower local trading volume and overlapping trading hours, and for stocks originating from the emerging markets.

In a traditional corporate finance area of dividend policy, Li *et al.* (2006) examine five determinants of dividend policy in Taiwanese companies. Their study employs a multinomial logistic model and divides the sample companies into two groups of high-tech and non-high-tech companies. They claim that the stability of dividends, debt ratio as well as profitability, are significant determinants of dividend policy regardless of high-tech and non-high-tech companies, and the size of the firm and its future growth opportunity have significant effects on dividend policy in non-high-tech companies but insignificant effects on high-tech companies. Readers with experience in financial analysis may recall the use of Z-scores in predicting bankruptcy by Altman (1968). Based on the analysis in this chapter, the probit model or the logit model should be superior in performing such a prediction. These non-linear models are attempted by Ginoglou *et al.* (2002) to predict corporate

failure of problematic firms in Greece, using a data set of 40 industrial firms, with 20 of them being healthy and 20 of them problematic, for the purpose of achieving improved prediction of corporate failure. However, the use of probit or logit models are not popular for classification issues involving accounting ratios, bankruptcy prediction in particular. Barniv and McDonald (1999) review several categorisation techniques and compare the performance of logit or probit with alternative procedures. The alternative techniques are applied to two research questions of bankruptcy prediction and auditors' consistency in judgements. Four empirical criteria provide some evidence that the exponential generalised beta of the second kind (EGB2), lomit, and burrit improve the log likelihood functions and the explanatory power, compared with logit and other models. EGB2, lomit and burrit also provide significantly better classifications and predictions than logit and other techniques.

   With aging becoming an increasingly important issue and current pension plans appear to be in crisis in many countries, research on pension or retirement plans is gaining momentum. In addition to Example 11.3 in the previous section, Cahill *et al.* (2005) have raised a question 'Are traditional retirements a thing of the past?' and attempted the question with new evidence on retirement patterns and bridge jobs. They investigate whether permanent, one-time retirements are coming to an end just as the trend towards earlier retirements did nearly 20 years ago. In particular, their study explores how common bridge jobs are among today's retirees, and how uncommon traditional retirements have become. The determinants of bridge jobs are examined with a multinomial logistic regression model, using data from the Health and Retirement Study for the work histories and retirement patterns of a cohort of retirees aged 51 to 61 in 1992 over a ten-year time period in both a cross-sectional and longitudinal context. They find that one-half to two-thirds of the respondents with full-time career jobs take on bridge jobs before exiting the labour force completely. Moreover, their study has documented that bridge job behaviour is most common among younger respondents, respondents without defined-benefit pension plans, and respondents at the lower- and upper-end of the wage distribution. They allege that traditional retirements will be the exception rather than the rule, based on the implications of their findings suggesting that changes in the retirement income landscape since the 1980s appear to be taking root.

   There have been many other applications of binary and discrete choice models in virtually every area of social sciences broadly defined. For example, Skaburskis (1999) adopts a multinomial logistic model and a naïve nested logistic model to study the choice of tenure and building type. These models are estimated to relate differences in household size, age, income and housing prices to differences in the choice of tenure and building type. The estimated models show that the demand for lower-density options decreases with a decrease in household size and an increase in age and real income. Higher rents tend to induce the demand for owner-occupied single-family detached houses. Higher price levels for the ownership option shift demand towards all of the higher-density options, particularly towards high-rise rental apartments. The effects of public premiums on children's health insurance

coverage are investigated by Kenney *et al.* (2006). They employ multinomial logistic models, focusing on premium costs and controlling for other factors. The study uses 2000 to 2004 Current Population Survey data to examine the effects of public premiums on the insurance coverage of children whose family income is between 100 per cent and 300 per cent of the federal poverty level. The magnitude of the estimated effects varies across the models. Nevertheless, it is claimed that the results consistently indicate that raising public premiums reduces enrolment in public programmes, with some children who forgo public coverage having private coverage instead and others being uninsured. The results further indicate that public premiums have larger effects when applied to lower-income families.

Questions and problems

1 Describe binary choice and, in general, discrete choice, with a few examples in daily life, corporate and individual.
2 What is defined as a limited dependent variable? What are features of limited dependent variables?
3 Describe and discuss the probit model with regard to its functional form.
4 Describe and discuss the logit model with regard to its functional form and in relation to logistic regression.
5 Contrast the probit model with the logit model, paying attention to their probability density functions.
6 Present and describe the multinomial logit model and multinomial logistic regression, and further discuss their role in modelling discrete choice.
7 What are defined as ordered probit and ordered logit? What differentiates an ordered logit model from a multinomial logit model?
8 What are marginal effects in discrete choice models? Why are the issues of marginal effects raised and addressed specifically for discrete choice models but not mentioned for linear regression and other linear models?
9 Collect data from various sources, e.g. *Acquisition Monthly*, Thomson ONE Banker and company annual reports, and then estimate a probit model and a logistic regression model for the choice of payment methods in mergers and acquisitions (using LIMDEP, GAUSS, RATS or other packages). The dependent variable is choice of payment methods, with choice of cash as 0 and choice of share exchange as 1. The independent variables may include the relative size of the bidder to the target, a measure of free cash flow of the bidder, a measure of the performance of the bidder and that of the target.
10 Collect data on dual listings from various sources, e.g. Thomson ONE Banker, the websites of relevant stock exchanges and companies, and then run multinomial logistic regression for the choice of foreign stock exchanges for dual listings (using LIMDEP, GAUSS, RATS or other packages). The dependent variable is choice of foreign stock exchanges for dual listings, with choice of New York being 0, London being 1 and Tokyo being 2 (or the reader's own choice of foreign stock exchanges). The independent variables may include firm size, a measure of performance, financial leverage, an

industry dummy, and a region dummy, or any variables that the reader reckons reasonable.

References

Altman, E.I. (1968), Financial ratios, discriminant analysis and prediction of corporate bankruptcy, *Journal of Finance*, **23**, 589–609.

Barniv, R. and McDonald, J.B. (1999), Review of categorical models for classification issues in accounting and finance, *Review of Quantitative Finance and Accounting*, **13**, 39–62.

Cahill, K.E., Giandrea, M.D. and Quinn, J.F. (2005), Are traditional retirements a thing of the past? new evidence on retirement patterns and bridge jobs, *U.S. Bureau of Labor Statistics Working Papers*: 384.

Chakravarty, S., Chiyachantana, C.N. and Jiang, C. (2004), The choice of trading venue and relative price impact of institutional trading: ADRs versus the underlying securities in their local markets, *Purdue University Economics Working Paper*.

Espahbodi, H. and Espahbodi, P. (2003), Binary choice models and corporate takeover, *Journal of Banking and Finance*, **27**, 549–574.

Ginoglou, D., Agorastos, K. and Hatzigagios, T. (2002), Predicting corporate failure of problematic firms in Greece with LPM logit probit and discriminant analysis models, *Journal of Financial Management and Analysis*, **15**, 1–15.

Howorth, C., Peel, M.J. and Wilson, N. (2003), An examination of the factors associated with bank switching in the U.K. small firm sector, *Small Business Economics*, **20**, 305–317.

Kenney, G., Hadley, J. and Blavin, F. (2006–2007), Effects of public premiums on children's health insurance coverage: Evidence from 1999 to 2003, *Inquiry*, Winter 2006–2007, **43**, 345–361.

Kim, H. and DeVaney, S.A. (2005), The selection of partial or full retirement by older workers, *Journal of Family and Economic Issues*, **26**, 371–394.

Koyuncu, C. and Bhattacharya, G. (2004), The impacts of quickness, price, payment risk, and delivery issues on on-line shopping, *Journal of Socio-Economics*, **33**, 241–251.

Li, M.Y.L., Wang, M.L., Wang, A.T. and Wang, C.A. (2006), Determinants of dividend policy of high-tech versus traditional companies: An empirical study using Taiwanese data, *Empirical Economics Letters*, **5**, 105–115.

Metwally, M.M. and Prasad, J. N. (2004), Factors restricting use of credit cards in GCC countries, *International Journal of Applied Business and Economic Research*, **2**, 171–188.

Parsons, L.M. (2007), The impact of financial information and voluntary disclosures on contributions to not-for-profit organizations, *Behavioral Research in Accounting*, **19**, 179–196.

Reiljan, E. (2002), Analysis of foreign direct investment determinants in Estonia, *Journal of East-West Business*, **8**, 103–121.

Skaburskis, A. (1999), Modelling the choice of tenure and building type, *Urban Studies*, **36**, 2199–2215.

12 Limited dependent variables and truncated and censored samples

In addition to discrete choice models where a dependent variable possesses discrete values, the values of dependent variables can also be censored or truncated. That is, the variable is not observed over its whole range. For example, in a survey of MBA graduates two years after their graduation, relevant information is collected, including their salaries. If one is interested in what determines the salary level, a model may be set up accordingly where salary is the dependent variable. The survey sets a lower limit and an upper limit for the salary range in case the respondents are not willing to reveal the exact salary beyond a certain range. Therefore, the value of the salary variable takes the figure given in the response when it falls in the given range, but is censored at both the lower and upper end. If a graduate does not respond, the observation corresponding to him is excluded or truncated from the sample, and neither the dependent variable nor the independent variables are available.

A dependent variable that is truncated or censored is a limited dependent variable, as is a dependent variable that is discrete discussed in the previous chapter. This chapter pays attention to issues in estimation of models involving limited dependent variables with regard to censored and truncated samples. Estimation of truncated or censored samples with conventional regression procedures such as the OLS can cause bias in parameter estimates. To correct for the bias, special techniques and treatments pertinent to truncated and censored data have to be applied. However, the problem can be largely solved by using the maximum likelihood estimation procedure. So, we present the likelihood and log likelihood functions for truncated and censored data samples first while demonstrating their distributions. Then we proceed to discuss the issue of bias in parameter estimation produced by the OLS procedure of the Tobit model, only for the purpose to introduce a wider issue of selection bias. Selection bias is addressed by referring to the work of Heckman (1976, 1979) and Cragg (1971).

12.1. Truncated and censored data analysis

We have learned limited dependent variables whose values are limited to discrete values in the previous chapter. There are other types of limited dependent variables that we introduce and study in this chapter. From time to time, samples can

be limited due to the values of the dependent variable. When this happens, the dependent variable is referred to as limited dependent variables. The two frequently experienced situations where samples are limited by limited dependent variables are truncated and censored samples. With a truncated sample, an observation is excluded from the sample if the value of the dependent variable does not meet certain criteria. In a censored sample, the dependent variable is not observed over its entire range for some of the observations.

A sample can be left truncated, right truncated, or truncated at both ends. In a left truncated sample, being truncated at TR_a, any observations are excluded from the sample if the dependent variable is smaller than or equal to TR_a:

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad Y_i > TR_a, \quad i = 1, \dots, N \qquad (12.1)$$

The probability of $Y_i > TR_a$ is the probability of $\varepsilon_i > TR_a - \mathbf{X}_i\boldsymbol{\beta}$, the cause of errors contributing to the bias, which should be subtracted from the likelihood function. Therefore, the likelihood function is adjusted and becomes:

$$\begin{aligned}
L &= \prod_{i=1}^{N} \phi\left(Y_i - \mathbf{X}_i\boldsymbol{\beta}\right)\left[1 - \Phi\left(TR_a - \mathbf{X}_i\boldsymbol{\beta}\right)\right] \\
&= \prod_{i=1}^{N} \frac{1}{\sigma}\phi\left(\frac{Y_i - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta} - TR_a}{\sigma}\right)
\end{aligned} \qquad (12.2)$$

where $\phi(\varepsilon)$ is the density function of normal distributions and $\Phi(\varepsilon)$ is the cumulative normal distribution, with $\varepsilon \sim N\left(0, \sigma^2\right)$. Note $\phi(\varepsilon/\sigma) = \sigma\phi(\varepsilon)$, $\Phi(\varepsilon/\sigma) = \Phi(\varepsilon)$ and $\Phi(-\varepsilon) = 1 - \Phi(\varepsilon)$. The corresponding log likelihood function following the adjustment is:

$$\begin{aligned}
LL &= \sum_{i=1}^{N} \text{Ln}\left[\phi\left(Y_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right] - \sum_{i=1}^{N} \text{Ln}\left[1 - \Phi\left(TR_a - \mathbf{X}_i\boldsymbol{\beta}\right)\right] \\
&= \frac{1}{\sigma}\sum_{i=1}^{N} \text{Ln}\left[\phi\left(\frac{Y_i - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right] - \sum_{i=1}^{N} \text{Ln}\left[\Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta} - TR_a}{\sigma}\right)\right] \\
&= -\frac{N}{2}\text{Ln}\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2 - \sum_{i=1}^{N} \text{Ln}\left[\Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta} - TR_a}{\sigma}\right)\right]
\end{aligned} \qquad (12.3)$$

For a right truncated sample, being truncated at TR_b, any observations are excluded from the sample if the dependent variable is greater than or equal to TR_b:

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i, \qquad \varepsilon_i \sim N\left(0, \sigma^2\right), \quad Y_i < TR_b, \quad i = 1, \dots, N \qquad (12.4)$$

Following similar considerations as for the left truncated sample, the log likelihood function of a right truncated sample is:

$$LL = \sum_{i=1}^{N} Ln\left[\phi\left(Y_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right] - \sum_{i=1}^{N} Ln\left[1 - \Phi\left(\mathbf{X}_i\boldsymbol{\beta} - TR_b\right)\right]$$

$$= \frac{1}{\sigma}\sum_{i=1}^{N} Ln\left[\phi\left(\frac{Y_i - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right] - \sum_{i=1}^{N} Ln\left[\Phi\left(\frac{TR_b - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right]$$

$$= -\frac{N}{2} Ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2 - \sum_{i=1}^{N} Ln\left[\Phi\left(\frac{TR_b - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right]$$

(12.5)

and the log likelihood function of a sample that is truncated at both ends is:

$$LL = \sum_{i=1}^{N} Ln\left[\phi\left(Y_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right]$$

$$- \sum_{i=1}^{N} Ln\left[1 - \Phi\left(TR_a - \mathbf{X}_i\boldsymbol{\beta}\right)\right] - \sum_{i=1}^{N} Ln\left[1 - \Phi\left(\mathbf{X}_i\boldsymbol{\beta} - TR_b\right)\right]$$

$$= \frac{1}{\sigma}\sum_{i=1}^{N} Ln\left[\phi\left(\frac{Y_i - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right]$$

$$- \sum_{i=1}^{N} Ln\left[\Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta} - TR_a}{\sigma}\right)\right] - \sum_{i=1}^{N} Ln\left[\Phi\left(\frac{TR_b - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right]$$

$$= -\frac{N}{2} Ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2$$

$$- \sum_{i=1}^{N} Ln\left[\Phi\left(\frac{TR_b - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right) + \Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta} - TR_a}{\sigma}\right)\right], \quad TR_a < TR_b \quad (12.6)$$

The presentation of the likelihood function for censored data is similar. Consider a sample that is left censored at c_a:

$$Y_i = \begin{cases} Y_i^* = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i, & \varepsilon_i \sim N\left(0, \sigma^2\right), & \text{if } Y_i > c_a \\ c_a, & & \text{if } Y_i \le c_a \end{cases} \quad i = 1, \dots, N \quad (12.7)$$

For non-censored observations, their likelihood function is:

$$L_i = \phi\left(Y_i - \mathbf{X}_i\boldsymbol{\beta}\right) = \frac{1}{\sigma}\phi\left(\frac{Y_i - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right) \tag{12.8}$$

and their log likelihood function is the standard log likelihood in conventional regression:

$$LL_i = \text{Ln}\left[\phi\left(Y_i - \mathbf{X}_i\boldsymbol{\beta}\right)\right] = -\frac{1}{2}\,\text{Ln}\left(2\pi\sigma^2\right) - \frac{(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2}{2\sigma^2} \tag{12.9}$$

For censored observations, their likelihood function is:

$$L_i = \Phi\left(c_a - \mathbf{X}_i\boldsymbol{\beta}\right) = \Phi\left(\frac{c_a - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right) \tag{12.10}$$

and their log likelihood function becomes:

$$LL_i = \text{Ln}\left[\Phi\left(\frac{c_a - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right] = \text{Ln}\left[\Phi\left(c_a - \mathbf{X}_i\boldsymbol{\beta}\right)\right] \tag{12.11}$$

So, the likelihood function of the sample is:

$$
\begin{aligned}
L &= \prod_{i=1}^{N} \phi\left(Y_i - \mathbf{X}_i\boldsymbol{\beta}\right)^{(1-\delta_i)} \Phi\left(c_a - \mathbf{X}_i\boldsymbol{\beta}\right)^{\delta_i} \\
&= \prod_{i=1}^{N} \left[\frac{1}{\sigma}\phi\left(\frac{Y_i - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right]^{(1-\delta_i)} \Phi\left(\frac{c_a - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)^{\delta_i}
\end{aligned}
\tag{12.12}
$$

where $\delta_i = 1$ if the observations are censored and $\delta_i = 0$ if the observations are not censored. The log likelihood function of the sample, therefore, is:

$$LL = -\sum_{i=1}^{N} (1-\delta_i)\left[\frac{1}{2}\text{Ln}\left(2\pi\sigma^2\right) + \frac{(Y_i - \mathbf{X}_i\boldsymbol{\beta})^2}{2\sigma^2}\right] + \sum_{i=1}^{N} \delta_i \text{Ln}\left[\Phi\left(\frac{c_a - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right] \tag{12.13}$$

For a sample that is right censored at c_b:

$$Y_i = \begin{cases} Y_i^* = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i, & \varepsilon_i \sim N\left(0, \sigma^2\right), \quad \text{if } Y_i < c_b \\ c_b, & \text{if } Y_i \geq c_b \end{cases} \quad i = 1, \ldots, N \tag{12.14}$$

the log likelihood function of non-censored observations is the standard log likelihood in conventional regression as presented in equation (12.9). The log likelihood function of censored observations is:

$$LL_i = \text{Ln}\left[1 - \Phi\left(\frac{c_b - \mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right] = \text{Ln}\left[\Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta} - c_b}{\sigma}\right)\right] \tag{12.15}$$

So, the log likelihood function of the sample is:

$$LL = -\sum_{i=1}^{N}(1-\vartheta_i)\left[\frac{1}{2}\mathrm{Ln}(2\pi\sigma^2)+\frac{(Y_i-\mathbf{X}_i\boldsymbol{\beta})^2}{2\sigma^2}\right]+\sum_{i=1}^{N}\vartheta_i\mathrm{Ln}\left[\Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta}-c_b}{\sigma}\right)\right]$$

$$(12.16)$$

where $\vartheta_i = 1$ if the observations are censored and $\vartheta_i = 0$ if the observations are not censored.

In general, a sample that is left censored at c_a and right censored at c_b has the following log likelihood function:

$$LL = -\sum_{i=1}^{N}(1-\delta_i-\vartheta_i)\left[\frac{1}{2}\mathrm{Ln}(2\pi\sigma^2)+\frac{(Y_i-\mathbf{X}_i\boldsymbol{\beta})^2}{2\sigma^2}\right]$$

$$+\sum_{i=1}^{N}\delta_i\mathrm{Ln}\left[\Phi\left(\frac{c_a-\mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)\right]+\sum_{i=1}^{N}\vartheta_i\mathrm{Ln}\left[\Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta}-c_b}{\sigma}\right)\right], \quad c_a < c_b$$

$$(12.17)$$

where $\delta_i = 1$ if the observations are left censored and $\delta_i = 0$ if the observations are not left censored, $\vartheta_i = 1$ if the observations are right censored and $\vartheta_i = 0$ if the observations are not right censored. Note any one observation cannot be left censored and right censored at the same time, i.e., $Y_i \le c_a$ and $Y_i \ge c_b$ cannot hold together. So the circumstance of $\delta_i = 1$ and $\vartheta_i = 1$ will not occur for any observation.

12.2. The Tobit model

The Tobit model is a kind of the modelling method for censored data samples, named after Tobin (1958). To simplify the matter, which is also fairly real practically, the model is usually left censored at 0, taking the form:

$$Y_i = \begin{cases} Y_i^* = \mathbf{X}_i\boldsymbol{\beta}+\varepsilon_i, & \varepsilon_i \sim N(0,\sigma^2), & \text{if } Y_i^* > 0 \\ 0, & & \text{if } Y_i^* \le 0 \end{cases} \quad i = 1,\ldots,N$$

$$(12.18)$$

Now consider a probit type model for the full sample, including both censored and non-censored data:

$$Y_i^* = \mathbf{X}_i\boldsymbol{\beta}+\varepsilon_i, \quad \varepsilon_i \sim N(0,\sigma^2)$$

$$(12.19a)$$

$$Y_i = \begin{cases} 1, & \text{if } Y_i^* > 0 \\ 0, & \text{if } Y_i^* \le 0 \end{cases}$$

$$(12.19b)$$

We have learned from the previous chapter that the likelihood function of the above probit model, $L(\boldsymbol{\beta})$, is:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{N} \left[\Phi(\mathbf{X}_i\boldsymbol{\beta})\right]^{Y_i} \left[1 - \Phi(\mathbf{X}_i\boldsymbol{\beta})\right]^{(1-Y_i)} \qquad (12.20)$$

The first moment or the expected value of Y_i^* conditional on $Y_i^* > 0$ is:

$$E(Y_i^* \mid Y_i^* > 0) = \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i \mid \varepsilon_i > -\mathbf{X}_i\boldsymbol{\beta})$$

$$= \mathbf{X}_i\boldsymbol{\beta} + \sigma E\left(\frac{\varepsilon_i}{\sigma} \mid \frac{\varepsilon_i}{\sigma} > -\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right)$$

$$= \mathbf{X}_i\boldsymbol{\beta} + \sigma \frac{\phi(\mathbf{X}_i\boldsymbol{\beta}/\sigma)}{\Phi(\mathbf{X}_i\boldsymbol{\beta}/\sigma)} = \mathbf{X}_i\boldsymbol{\beta} + \sigma \lambda_i\left(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma}\right) \qquad (12.21)$$

where λ_i is called the inverse Mill's ratio. Since this would be the estimates by the OLS including non-censored observations only, and is not equal to $\mathbf{X}_i\boldsymbol{\beta}$, the OLS is inconsistent and produces biased results. Note that the exclusion of the censored observations in the analysis amounts to estimation of a truncated example. That is, we run the OLS regression of the dependent variable on the explanatory variables with non-censored observations, the estimates are biased due to the restriction imposed on the sample or the exclusion of the censored observations. However, this bias may be corrected, since the bias arises from the given positive correlation between ε_i and $\mathbf{X}_i\boldsymbol{\beta}$; it is linked to the inverse Mill's ratio and, as a result, may be corrected. This analysis of bias suggests a two-step or two-stage procedure. In the first stage, a probit type model is estimated for the full sample, including both censored and non-censored data. The values of the dependent variable are made discrete; the values of the dependent variable for uncensored observations are set to 1 and the values of the dependent variable for censored observations are set to 0. The inverse Mill's ratio can be obtained accordingly. In the second stage of the estimation procedure, only non-censored data, or the truncated sample, are applied to the regression. The dependent variable is regressed on the explanatory variables and the inverse Mill's ratio obtained in the first stage estimation. This corrects the bias caused by the restriction imposed on sample selection or the exclusion of censored observations. This is the idea of the Heckman (1979) model, which generalises the Tobit model.

Let us review what is defined as the inverse Mill's ratio prior to proceeding to more general selection models in the next section, where this ratio is frequently used. Consider the probability density function of a standard normal distribution:

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2} \qquad (12.22)$$

Its first moment conditional on $z > c$:

$$E(z \mid z > c) = \int_c^\infty z\phi(z \mid z > c)\,\mathrm{d}z$$

$$= \int_c^\infty z\frac{\phi(z)}{P(z > c)}\,\mathrm{d}z = \frac{1}{1 - \Phi(c)}\int_c^\infty z\phi(z)\,\mathrm{d}z$$

$$= \frac{-1}{1 - \Phi(c)}\int_c^\infty \mathrm{d}[\phi(z)] = \frac{\phi(c)}{1 - \Phi(c)} \tag{12.23}$$

is known as the inverse Mill's ratio:

$$\lambda(c) = \frac{\phi(c)}{1 - \Phi(c)} = \frac{\phi(-c)}{\Phi(-c)} \tag{12.24}$$

In general, the inverse Mill's ratio for a normal distribution with a mean of μ and a standard deviation of σ, $N(\mu, \sigma^2)$, is:

$$\lambda(c) = \frac{\phi((c - \mu)/\sigma)}{1 - \Phi((c - \mu)/\sigma)} = \frac{\phi((\mu - c)/\sigma)}{\Phi((\mu - c)/\sigma)} \tag{12.25}$$

When $c = 0$, the above inverse Mill's ratio becomes:

$$\lambda(c) = \frac{\phi(\mu/\sigma)}{\Phi(\mu/\sigma)} \tag{12.26}$$

It is simply the ratio of the value of the probability density function to the value of the cumulative normal distribution at point μ.

The Tobit model is widely used for its simplicity, straightforwardness and relevance, despite the criticisms. More advanced models have been developed and extended, ranging from estimation procedures, such as the maximum likelihood, to model specifications, such as the Cragg (1971) model and the Heckman (1979) model to be introduced in the next section. Although the maximum likelihood procedure can produce consistent unbiased coefficient estimates, the two-stage OLS procedure is still used for the following reasons. Firstly, the bias can be easily corrected. The execution of OLS procedures is much simpler than the maximisation of the log likelihood function. The latter involves non-linear iterations, which is sensitive to the choice of initial values of the parameters and may fail to converge sometimes, despite that computing time and speed are no longer a concern nowadays. Secondly, we would like to know the different effects of the dependent variable on parameter estimates by using the truncated sample and the 'full' sample that does not exist, or the extent to which the estimated parameter is biased. Some inference for the 'full' sample's characteristics may result from such analysis.

12.3. Generalisation of the Tobit model: Heckman and Cragg

This section generalises the Tobit model. The basic Tobit model introduced in the previous section, which can be called the Tobit type I model, assumes that the two processes underlying the continuous choice and discrete choice are the same. The two processes have the same independent variables to explain the probability of the discrete dependent variable being observed, as well as the magnitude of the continuous dependent variable. However, the two processes may be different involving different independent variables. For example, the decision to participate in an investment activity may not totally be based on factors that determine the level or extent of involvement in that activity by those who have participated in. The Heckman model, which is also referred to as the Tobit type II model, generalises the Tobit type I model by modelling the decision process and the level of involvement with two different processes. That is, the set of the variables that influence the decision to participate may be different from, albeit could be identical to, the set of variables that determine the level or extent of participation. This decision process 'selects' observations to be observed in the truncated data sample that consists of non-censored data only. When the set of variables for the decision equation is the same as that for the level or extent equation, the model reduces to the basic Tobit model, or the Tobit type I model, where observed observations are non-censored and the rest are censored in a censored data sample, and the selection process truncates the sample by selecting non-censored data in the regression. There would be selection bias in the process, which should be detected if exists, and be corrected accordingly, usually by the means of applying the inverse Mill's ratio.

A typical Heckman's model has the following representation:

$$\left.\begin{aligned} Y_i^* &= \mathbf{X}_i\boldsymbol{\beta} + \mu_i \\ Z_i^* &= \mathbf{W}_i\boldsymbol{\gamma} + v_i \end{aligned}\right\} \tag{12.27a}$$

$$\begin{pmatrix} \mu_i \\ v_i \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\mu^2 & \sigma_{\mu v} \\ \sigma_{v\mu} & \sigma_v^2 \end{bmatrix}\right), \quad i = 1, \dots, N$$

$$Y_i = \begin{cases} Y_i^*, & \text{if } Z_i^* > 0 \\ 0, & \text{if } Z_i^* \leq 0 \end{cases} \quad i = 1, \dots, N \tag{12.27b}$$

The purpose is to estimate the Y_i^* equation, but certain observations of Y_i^* are not available. The inclusion of Y_i^* in the sub-sample of available observations is determined by a selection rule, a process described by the Z_i^* equation. Since the dataset is incomplete, or only the data in a selected sub-sample are observed, the OLS regression for the sub-sample may differ from the OLS regression for the whole sample in the Y_i^* equation. The difference between the two sets of parameter estimates based on the sub-sample and the whole sample arises from the selection rule or selection process and, is therefore referred to as selection bias. The selection bias can be illustrated as follows when estimation of the conditional

mean of Y_i^*, the mean of Y_i^* conditional on $Z_i^* > 0$, is attempted:

$$
\begin{aligned}
E(Y_i^* \mid Z_i^* > 0) &= E(Y_i^* \mid v_i > -\mathbf{W}_i\boldsymbol{\gamma}) \\
&= \mathbf{X}_i\boldsymbol{\beta} + E(\mu_i \mid v_i > -\mathbf{W}_i\boldsymbol{\gamma}) \\
&= \mathbf{X}_i\boldsymbol{\beta} + \frac{\sigma_{\mu v}}{\sigma_v}\, \frac{\phi\left(Z_i\gamma/\sigma_v\right)}{\Phi\left(Z_i\gamma/\sigma_v\right)} \\
&= \mathbf{X}_i\boldsymbol{\beta} + \frac{\sigma_{\mu v}}{\sigma_v}\lambda_i
\end{aligned}
\tag{12.28}
$$

where λ_i is the inverse Mill's ratio. In recognition of this selection bias, Heckman proposes a two-stage procedure to obtain consistent parameter estimates. In the first step, the parameters of the probability of $Z_i^* > 0$ are estimated using a probit model for the whole sample. The inverse Mill's ratio is estimated during the course. All these parameter estimates are consistent. In the second step, the following equation is estimated by the OLS:

$$
Y_i^* = \mathbf{X}_i\boldsymbol{\beta} + \xi\lambda_i + \varepsilon_i
\tag{12.29}
$$

for the sub-sample selected by $Z_i^* > 0$. The resulting parameter estimates are consistent. If ξ is significantly different from zero, i.e., if the hypothesis $\xi = 0$ is rejected, then there exists selection bias. Otherwise, there is no selection bias, or the Z_i^* equation is irrelevant and can be dropped out from the analysis.

12.4. Examples and cases

Example 12.1

In a study of bank expansions in California, Ruffer and Holcomb Jr (2001) use a typical two stage estimation procedure, examining the expansion decision by banks first, and then the level of expansions or the expansion activity rate next. A probit model is employed for the first stage analysis of expansion decisions, which also produces the inverse Mill's ratio for the second stage analysis to correct selection biases in the examination of expansion activity rates in a Heckman type regression. The authors choose the same variables for both the decision model and the level model except the inclusion of a new market dummy in the second stage analysis, and justify their stance by stating that they feel there are no theoretical reasons to exclude any of the variables in either stage of the estimation. The variables that are considered to be relevant to expansion decisions and activity levels

have been chosen as future population growth, market deposit, bank deposit, market share, the Herfindahl Hirschman index as a measure for market concentration, and bank specialisation. The new market dummy variable is excluded from the probit model because, by definition, a bank that expands in a new market is engaged in an expansion.

There are two types of expansions in this study: expansion by acquiring a bank branch (to buy) and expansion by building a new branch (to build). The authors examine these two types of expansions separately in the second stage analysis, though the two types of expansions are aggregated in the probit model for expansion decisions. The data set used in this study consists of 70,246 observations. There are 65,763 observations where the bank is not present in a particular market, so an expansion decision is to expand into a new market for these observations. In most cases reflecting these observations, banks choose not to expand. Only 71 out of the 65,763 observations correspond to an expansion where the bank does not exist at the beginning of the period. In their probit model, the bank makes the expansion decision for each market and time period. Since there are 31 relevant geographic markets in the data set, each bank in existence in a given year has 31 observations, one for each market. There are 517 different banks and 5,450 different branches covered by the study. The number of branches per bank ranges from a minimum of one to a maximum of 1,007. Over the five year period covered by the data set, 590 branches are built, 629 branches are acquired, and 38 banks are involved in an acquisition. Of the 590 branches that are built, 69 cases are expansions into a new market; and of the 629 acquisitions, only two of them expand into a new market. Since some banks have acquired more than one bank branch in one year, the number of acquisitions is not the same as the number of banks involved in acquisitions.

The results from estimating the decision model and the level model are reported in Table 12.1. Four year dummies are incorporated to catch the difference in expansion activities taking place in different years, as the data are not arranged as a panel data set, which is to be introduced in the next chapter. The results indicate that market deposits, bank deposits, and market share have a positive effect on bank expansion decisions, while market concentration, measured by the Herfindahl Hirschman index, is negatively associated with bank expansion decisions. That is, when a market becomes larger in terms of total market deposits in that market, probability of expanding in that market by a bank becomes higher. Larger banks, with larger bank deposits, are more likely to engage in expansion activities. As a bank's market share becomes greater, the bank is more likely to expand. The authors claim that market concentration is a barrier to entry, based on

Continued

Table 12.1 Decision model of expansion and level models of modes of expansion

	Probit	Acquision activity rate	Building activity rate
Constant	−2.1769	−46.4050	−37.1714
	(.1135)	(30.9878)	(24.8817)
Future population growth	−2.01132	−0.3659	−.2453
	(.006829)	(.4574)	(0.3897)
Market deposit	.8781e−8***	0.1315e−6	0.2176e−6**
	(.9056e−9)	(0.1068e−6)	(.8695e−7)
Herfindahl Hirschman index	−2.2904***	−75.0874**	19.7346
	(0.3882)	(34.0893)	(28.2156)
Bank deposit	.2516e−7***	0.7744e−6***	.1110e−6
	(.4044e−8)	(0.2768e−6)	(.2211e−6)
Market share	3.7910***	63.2080	18.4181
	(.6109)	(46.9135)	(37.9271)
Bank specialisation	.9925***	17.4543	7.3618
	(.05279)	(9.3844)	(7.4099)
Year 1 (1985 dummy)	−.04203	−9.0733**	4.6661
	(.05816)	(3.5729)	(3.0358)
Year 2 (1986 dummy)	−.02342	1.0285	−.5466
	(0.05725)	(3.4595)	(2.9419)
Year 3 (1987 dummy)	0.06288	−.5245	−1.1592
	(.05313)	(3.314)	(2.8162)
Year 4 (1988 dummy)	−.03432	−7.0231**	.8277
	(0.05466)	(3.3087)	(2.8083)
New market		17.2656***	58.1125***
		(3.00027)	(2.6511)
Lambda		25.1645**	16.9925
		(12.1623)	(9.7621)
R^2		0.1208	0.5834

Standard errors in parentheses.
*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.

the negative and statistically significant coefficient for the Herfindahl Hirschman index. In addition, they allege that the role of bank specialisation is as conjectured: a positive and statistically significant coefficient implies that a bank is more likely to expand in a market in which it has a larger percentage of its current total deposits invested. Nevertheless, while the coefficient is able to indicate expansion, it seems to be unable to identify the market in which a bank is to expand. Population growth is found to have no effect on expansion decisions, and there are no year to year variations as indicated by the insignificant year dummies.

Although the authors point out that they use the same variables for both the decision model and the level model, the 'same' variables in their models are in fact not the same. For example, the bank deposit data of the banks that choose to expand in the decision model cover the data for both modes of expansion by acquiring an existing bank branch and building a new branch; whereas the level model deals with the expansion activities separately for each of the two modes of expansion. Yet, the inverse Mill's ratio from running the probit model is valid theoretically and helpful to the second stage OLS estimation of the two level models. The level model results appear to be reasonable that the market wide measure of market deposits contribute positively to building activity rates while the bank specific measure of bank deposits contribute positively to acquisition activity rates. When a market becomes larger, fewer bank branches are for sale in that market, and the feasible way of expansion seems to be building new branches; whereas larger banks are more inclined to engage in acquisition activities positively. Obviously, the Herfindahl Hirschman index plays a different role in the two modes of expansion. The coefficient for the index is negatively significant in the level model for acquisition activity rates, and it is positive but insignificant in the level model for building activity rates. When a market is more concentrated, fewer acquisitions take place, resulting in a negatively significant coefficient. Expansions in a more concentrated market tend to be building new branches, albeit this mode of expansion is also unlikely to be attempted when market concentration goes up, suggested by the fact that the coefficient, though positive, is insignificant. The authors concede that some of the variables that play an important role in the probit model do not have significant effect on the level of expansion activities, which may question the selection of variables, especially for the second stage regression. For instance, the coefficient for the market share variable and that for the bank specialisation variable are insignificant in both cases of acquisition activity rates and building activity rates. Future population growth has no effect on the level of expansion as well as the decision to expand, i.e., the variable plays no role at all in this study. Acquisition activities are significantly lower in 1985 and 1988 than in the benchmark year of 1989, though the background to the drop in acquisition activities is not explicated. The study also divides the sample into two sub-samples according to bank size, and repeats the above estimation procedure for small banks and large banks respectively. It is reported that the results for small banks are similar to those for the whole sample. For large banks, only market deposits and bank deposits have a positive effect on the decision to expand as expected. None of the variables have any effect on the level of expansion activities, except the new market dummy that is positive and significant, indicating that expansions into new markets by large banks are primarily through the building of new branches in the case of this study.

Example 12.2

This study by Przeworski and Vreeland (2000) examines the effect of IMF programmes on participating countries' economic growth, which adopts Heckman's procedure to correct the selection bias in their estimation of economic growth. The first stage analysis involves probit decision models for entering an IMF programme and remaining under an IMF programme. Participation in an IMF programme is a joint decision by the government and the IMF. Therefore, for a country to enter and to remain under IMF agreements, both the government and the IMF must want to do so for varied reasons. The authors have identified seven variables for the decision model for entering and remaining under the IMF programme by governments. They are reserves, deficit, debt service, investment, years under, number under, and lagged election. The reserve variable is the average annual foreign reserves in terms of monthly imports. Deficit, measured as budget surplus following the World Bank convention, is the annual government budget surplus as a proportion of GDP. Debt service is the annual debt service as a proportion of GDP. Investment is real gross domestic investment as a proportion of GDP. Years under, the number of years a country has spent under IMF agreements, is considered to be relevant to governments' IMF participation decisions. Number under, the number of other countries around the world currently participating in IMF programmes, is also deemed to have an influence on a government's IMF participation decisions. Lagged election is a dummy variable taking the value of 1 if there was a legislative election in the previous year and 0 otherwise. The final data set used in this study covers 1,024 annual observations of 79 countries for the period between 1970 and 1990, after removing observations and countries with incomplete information from their basic data set that includes 4,126 annual observations of 135 countries for the period between 1951 and 1990.

The results from estimating the decision model by governments are reported in the upper panel of Table 12.2. All of the seven variables have a statistically significant effect on a government's decision to enter into an IMF programme; whereas only one of the variables, the number of other countries around the world currently participating in IMF programmes, influences a government's decision to remain under IMF agreements significantly at the 5 per cent level. The coefficient for the reserves variable is negative, being significant at the 5 per cent level. This is reasonable that governments are more likely to enter into IMF programmes when their countries' foreign reserves are low. Similar is the role of the deficit variable measured in budget surplus and of the investment variable. The coefficient for both variables is significantly negative at the 1 per cent level. That is, governments with high deficit or lower budget surplus are more likely to seek entry into IMF

Table 12.2 Decision models to enter into and remain under IMF programmes

	Decision model for entering		Decision model for remaining	
	Government			
Constant	−2.27***	(0.611)	−0.01	(0.592)
Reserves	−0.83*	(0.424)	−0.26	(0.475)
Deficit	−0.95***	(0.277)	−0.29	(0.329)
Debt service	1.38***	(0.516)	0.65	(0.678)
Investment	−6.06***	(1.789)	−0.17	(1.922)
Years under	0.36*	(0.212)	−0.36	(0.266)
Number under	0.44***	(0.178)	0.38	(0.190)
Lagged election	0.87***	(0.288)	−0.01	(0.352)
	IMF			
Constant	2.14*	(1.241)	2.84	(2.016)
BOP	−0.91***	(0.370)	−0.41*	(0.230)
Number under	−0.73***	(0.268)	−0.39	(0.429)
Regime	0.43*	(0.260)	0.33	(0.273)

Standard errors in parentheses.
*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.

programmes; and countries with lower real gross domestic investment are more inclined to enter as well. As expected, governments are more likely to enter into IFM agreements when debt service is higher, indicated by a positive coefficient that is significant at the 1 per cent level. The coefficient for the variable years under, the number of previous years a country has spent under IMF agreements, is positive but significant at a modest 10 per cent level. This indicates that a government is more inclined to enter into a new IMF agreement if the country has spent a longer period under IMF programmes in the past, though not resolutely. It appears that there is a peer effect on a government's decision to enter into IMF agreements, suggested by a positive coefficient, highly significant at the 1 per cent level, for the variable standing for the number of other countries around the world currently participating in IMF programmes. The lagged election dummy is also significant at the 1 per cent level, indicating that a government is more likely to sign an IMF agreement at a time when the next election is distant. Only one out of the seven variables, the number of other countries around the world currently participating in IMF programmes, has a positive peer effect on a government's decision to remain under IMF agreements. The coefficient for the variable is positive and significant

Continued

at the 5 per cent level, suggesting that a government is more likely to continue its IMF participation when more countries choose to involve IMF engagements.

The authors include three variables in the decision models for the IMF. They are BOP, the balance of payments; number under, the number of other countries around the world currently participating in IMF programmes, as a proxy for the IMF budget; and regime, a dummy variable taking the value of 1 for dictatorship governments and 0 for democracies. The results from estimating the decision model by the IMF are reported in the lower panel of Table 12.2. According to these results, the IMF is less likely to provide a country with a programme arrangement if the country has a larger balance of payments surplus, and vice versa, shown by a significantly negative coefficient for the variable at the 1 per cent level. The variable standing for the number of other countries around the world currently participating in IMF programmes is a proxy for the IMF budget here. A highly significantly negative coefficient for the variable indicates that the IMF acts under budget constraints – the IMF signs less countries when more countries are already under IMF programmes. The regime dummy is positive but only modestly significant at the 10 per cent level, suggesting the IMF is more likely to sign with dictatorship governments than democratic governments. The authors explain that this may be because dictatorship governments are easier to negotiate with; while in democracies oppositions may have more power to scrutinise the agreement with the IMF, so the initiated agreements may not be signed in the end. Regarding the decision of the IMF to retain countries under its programmes, only the balance of payments has a significant effect with the coefficient for the variable being negative and significant at the 10 per cent level. The authors allege that once negotiations have been concluded, the IMF's costs of negotiation have been met, and all that matters is whether balance of payments deficit continues to be large.

It is found that IMF programme participation lowers growth rates for as long as countries remain under a programme (Table 12.3). Once participating countries leave the programme, they grow faster than if they had remained under the programme, but not faster than they would have grown without IMF participation.

The second stage analysis of the study applies Heckman's selection bias correction procedure to the growth model, including the inverse Mill's ratios derived from the decision model as additional independent variables in the regression. The purpose is to evaluate the effect of IMF participation on growth. There were 465 observations for countries participating in IMF programmes; the observed average growth rate for these countries was 2.04 per cent per annum. There were 559 observations for countries not

Table 12.3 IMF programme participation and growth

	Under		Not under	
	Estimated coefficient	*Observed mean*	*Estimated coefficient*	*Observed mean*
Constant	-1.73^{***}	1.00	-0.13	1.00
	(0.44)		(0.38)	
$\dot{K}/K$	0.47^{***}	2.01	0.44^{***}	7.15
	(0.01)		(0.02)	
$\dot{L}/L$	0.53^{***}	2.80	0.56^{***}	2.69
	(0.01)		(0.02)	
λ^{G}	4.31^{***}	0.14	0.07	-0.73
	(1.48)		(0.23)	
λ^{I}	6.17^{***}	0.12	0.09	-0.82
	(2.23)		(0.29)	
$E\left(\dot{Y}/Y\right)$	2.00	2.04	3.53	4.39
	(5.93)	6.68	(5.50)	7.15
Observations	465		559	
Durbin–Watson	1.75		1.89	
Adjusted R^2	0.71		0.59	
F-test	0.00, $p = 0.99$		1.68, $p = 0.19$	

$\dot{k}/k$, growth in capital; $\dot{L}/L$, growth in labour; $E(\dot{Y}/Y)$, expected growth in output; λ^{G}, coefficient of the inverse Mill's ratio for government; λ^{I}, coefficient of the inverse Mill's ratio for IFM.
Standard errors in parentheses.
*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.

under IMF programmes; the observed average growth rate for these countries was 4.39 per cent per annum. The weighted average growth rate for participating and non-participating countries combined was 3.32 per cent per annum. The authors raise the question of whether this difference is due to the conditions the countries faced or due to IMF programme participation. They classify these conditions by the size of the domestic deficit and of foreign reserves. Scrutiny reveals that those countries under good conditions but not under IMF programmes were 1.02 per cent faster in growth than countries which experienced the same conditions while under IMF programmes; Those countries under bad conditions and under IMF programmes could have been 1.79 per cent faster in growth if they had not participated in IFM programmes. Accordingly they conclude that, while countries facing bad conditions grew slower, participation in IMF programmes lowered growth under all conditions.

12.5. Empirical literature

As has been introduced at the beginning of this chapter from time to time we encounter samples where data are truncated or censored. While we may not be able to get hold of a sample free of truncation or censoring under certain circumstances, correction for the bias introduced by truncation and censoring is a way to obtain correct parameter estimates. We will learn in the following that many studies adopt the strategy of bias correction modelling in empirical investigations into various issues in finance, economics and related areas. The techniques and special treatments pertinent to truncated and censored data discussed earlier in this chapter have been widely applied. In particular, the Heckman procedure is followed to highlight selection bias; whilst the maximum likelihood method is available that does not produce bias in estimation.

Diversification is at the centre of finance. While diversification in financial market investment is almost universally advocated, whether firms should diversify in their business lines has been subjected to intense debate. One of the most cited reasons is that, since shareholders can achieve diversification through their investment on the stock market, there is no need for the firm to do the job for them. Moreover, some have argued that, instead of generating any benefit from such firm-level diversification, the strategy actually destroys firm value. In a recent study, Villalonga (2004) re-examines the issue and raises the question 'does diversification cause the "diversification discount"'? The sample used in the study is gathered from Compustat and consists of 60,930 firm-years during 1978–1997, of which 20,173 firms-years are diversified or multi-segment and 40,757 are single-segment. The results from the preliminary analysis of cross-section discounts and longitudinal discounts indicate that diversification destroys value, in line with the results in the two studies the author refers to. Then, the study embarks on a two-stage estimation procedure. It uses a probit model to predict the propensity to diversify by firms. The independent variables employed at this stage include total assets, profitability, investment, percentage of ordinary shares owned by institutions, percentage of ordinary shares owned by insiders, R&D expenditure and firm age for firm characteristics; previous year's industry Q, fraction of firms in the industry that are diversified, and fraction of industry sales accounted for by diversified firms for industry characteristics; and a few variables for macroeconomic characteristics. In addition, dummies are used for membership of S&P, major exchange listings, foreign incorporation and dividend payouts, where the dummy takes the value of 1 if a firm belongs to S&P, listed on NYSE, AMEX or NASDAQ, is incorporated abroad, and paid dividends. The second stage analysis is to evaluate the effect of diversification on firm value through estimating 'the average treatment effect on the treated' for diversifying firms. The 'cause' is the treatment and the 'effect' is the change in excess value from year $t-1$ to year $t+1$. The treatment variable or indicator takes the value of 1 for firm-years in which the number of segments increases from one to two or more through either acquisitions or internal growth, and 0 for single-segment years. Three measures of excess value are adopted in the second stage analysis; they are

asset multiplier, sales multiplier and industry-adjusted Q. The change in excess value is the dependent variable in the value equation or second stage regression. The treatment indicator or variable is the principal independent variable in the second stage regression, into which the inverse Mill's ratio estimated from the first stage probit model enters as an additional variable in the Heckman model. Total assets, profitability, investment, previous year's industry Q and S&P dummy are included as control variables. If diversification destroys value, the coefficient of the treatment variable should be significantly negative for diversification to contribute to excess value negatively significantly. The author employs other models in addition to the Heckman's, and the results indicate the coefficients of the treatment variable are insignificant for all three measures of excess value estimated by all these models, though their signs are negative. Based on these results, the author claims that, on average, diversification does not destroy value.

One of the major phenomena in globalisation is increased activities in foreign direct investment (FDI) across borders, in addition to proliferating global portfolio investment. Firms from developed economies invest directly between them, in the developing world, and vice versa. In this area, the spillover effect from FDI has been scrutinised by Chuang and Lin (1999), Sinani and Meyer (2004) and Kneller and Pisu (2007). Sinani and Meyer (2004) conduct an empirical study to estimate the impact of technology transfer from FDI on the growth of sales of domestic firms in Estonia during the period from 1994 to 1999. The sample used in the study contains yearly information on Estonian firms from 1994 to 1999, obtained from the Estonian Statistical Agency. It consists of 2,250 observations, of which there are 405 firms in 1994, 434 in 1995, 420 in 1996, 377 in 1997, 320 in 1998 and 294 firms in 1999. After the first differencing operation being applied to sales and input variables, 1,339 observations remain for domestic firms and 359 for foreign firms. The authors adopt the Heckman two-stage procedure to control for sample self selection bias. Firstly, they estimate the probability that a firm is included in the sample based on the firm's profit, its labour productivity and its industry affiliation. Then, the resulting inverse Mill's ratio is included as an additional independent variable in the regression for spillovers effect to correct for selection bias or exclusions. The dependent variable is firm-level growth of sales for domestic firms. The principal independent variable is a measure of spillovers in the previous year. In the study, the share of foreign firms in industry employment, that in sales and that in equity are used as the measures of spillovers. Several control variables enter the regression as well, including intangible assets, investment in the previous year, human capital, the products of a previous year's measure of spillovers and intangible assets, investment in the previous year and human capital, export, the foreign Herfindahl Hirschman index, the domestic Herfindahl Hirschman index, industry dummies and time dummies. It is found that the magnitude of the spillover effect depends on the characteristics of incoming FDI and of the recipient local firm. More specifically, spillovers vary with the measure of foreign presence used and are influenced by the recipient firm's size, its ownership structure and its trade orientation. Chuang and Lin (1999) investigate the effect of FDI and R&D on productivity. Their data set includes 13,300 Taiwanese manufacturing firms

randomly sampled from, and accounted for about 9 per cent of all registered firms in the region, drawn from a report on 1991 industrial and commercial consensus. The final sample consists of 8,846 firms after removing firms with incomplete data on certain variables. Part of their paper is on the relationship between FDI and R&D. They claim that firms may self-select into R&D or non-R&D groups, and apply Heckman's two-stage procedure, which confirm the existence of selection bias with a significant inverse Mill's ratio in the second stage regression. In the first stage probit analysis, the firm's tendency or likelihood of engaging in R&D activities is modelled in such a way that it may be influenced by the following variables: FDI in terms of the share of foreign assets at the industry level, a dummy variable representing foreign ownership of the firm, a dummy variable representing technology purchase, outward foreign investment, the capital-labour ratio, and the age of the firm since its establishment. All these independent variables are found to have a positive effect on firms' likelihood of engaging in R&D activities at the 5 per cent significance level statistically. In the second stage regression, the authors employ the same variables to estimate their effect on R&D expenditure, measured by the ratio of the firm's R&D expenditure to its total sales. When the inverse Mill's ratio is not included in the regression, the FDI variable is found to have a positive effect on R&D expenditure. However, the inverse Mill's ratio, when included in the regression, is significant at the 5 per cent level, indicating that firms self-select into the R&D group. With the inverse Mill's ratio being included, the coefficient of the FDI variable becomes insignificant with a negative sign, which implies that there is no clear relationship between FDI and R&D activities. The authors interpret their results by referring to the cited studies in their paper that local firms need to enhance their technical capacity via their own R&D first, in order to absorb and digest new technology from abroad. So, FDI and R&D tend to be complementary. The rest of the variables that have an effect on R&D expenditure are foreign ownership, technology purchase, age and outward foreign investment. In contrast to the probit results, the effect on the level of R&D activity of these variables is all negative, to which the authors have offered some explanations. In a recent paper, Kneller and Pisu (2007) study export spillovers from FDI. In particular, their empirical study is centred on two equations of export decision and export share regression, where the independent variables entering both equations are the same. The coefficient of the inverse Mill's ratio is reported in the paper to indicate the existence of selection bias. The authors obtain their data from OneSource: UK Companies Volume 1, October 2000; and the number of observations included in the study are 19,066 in the time period of 1992–1999. The effect of foreign presence is estimated, using the complied horizontal measurement, the backward measurement and the forward measurement indices as independent variables. According to and interpreting the results, the decision of domestic firms on export engagement does not seem to be influenced by contacts with multinational corporations. However, foreign presence in the same upstream and downstream industries is found to have an effect on the amount of export by firms. Significantly negative forward export spillovers and significantly positive backward spillovers are found to exist. Export-oriented multinationals produce

positive horizontal export spillovers statistically significantly; whereas the same effect is also found from domestic market-oriented multinationals, albeit to a less extent.

Testing Gibrat's law for small, young and innovating firms, Calvo (2006) investigates whether small, young and innovating firms have experienced a greater employment growth than other Spanish firms over the period 1990–2000. The sample used in the study consists of 1,272 manufacturing firms, 967 of them survived for the entire ten year period, which results in selection bias. Therefore, Heckman's two-stage procedure is adopted to correct the selection bias, where the inverse Mill's ratio obtained from the first stage probit model is incorporated in the second stage regression. In addition, the maximum likelihood method is employed. The author alleges that all his results reject Gibrat's law and support the proposition of the paper that small firms have grown larger; moreover, old firms grow less than young ones. Innovation in process and product is found to have a significant positive effect on the firm's survival and its employment growth. Power and Reid (2005) examine the relationship between flexibility, firm-specific turbulence and the performance of long-lived small firms. Their study is fieldwork based and uses information gathered directly from small firm entrepreneurs. Measures of flexibility and firm-specific turbulence with 28 distinct attributes are used in the study. Then the effect of flexibility and firm-specific turbulence on long-run performance is estimated, using GLS and Heckman's selectivity correction procedure. It is claimed that firm-specific turbulence has a negative effect on performance, while greater flexibility of the small firm improves performance.

A great portion of studies that involve the problem of selectivity are in the areas of labour economics and human resource management, as they are where the issue originates. This section only selects two of them to illustrate the application, focusing on the financial aspect of the study. Gender and ethnic wage structures and wage differentials in Israel are examined by Neuman and Oaxaca (2005). They decompose the difference in wages into endowments, discrimination and selectivity components. Selection and wage equations are estimated for each of the demographic groups of Eastern women, Western women, Eastern men and Western men respectively. Heckman's two-stage procedure is used in the wage equations to correct selection bias. Then wage differentials are decomposed into that of endowments, discrimination and selectivity. Gender wage differentials are claimed to be significantly larger than ethnic differentials. Their four alternative decompositions yield different results. Information on the relative shares of the endowments, discrimination and selectivity components leads to a more effective way to close wage gaps. McCausland *et al.* (2005) investigate whether significant differences exist in job satisfaction (JS) between individuals receiving performance-related pay (PRP) and those on alternative compensation plans. The study uses data from four waves of the British Household Panel Survey (BHPS) 1998–2001, a nationally representative survey that interviews a random sample of around 10,000 individuals in approximately 5,500 households in Britain each year. It contains information on employees' personal and employment characteristics. Respondents in employment are asked about their satisfaction in seven aspects

of their jobs, i.e., promotion prospects, total pay, relations with supervisors, job security, ability to work on own initiative, the actual work itself and hours of work. They are also asked to rate their overall JS. The final sample used in the study consists of 26,585 observations on 9,831 individuals, 16.26 per cent of them are with PRP schemes. This sample is split into PRP and non-PRP sub-samples in empirical analysis. Heckman's two-stage procedure is then applied to correct for both self-selection of individuals into their preferred compensation scheme and the endogeneity of wages in a JS framework. A probit model is estimated in the first stage analysis. The resulting inverse Mill's ratios are then included in the second stage regression equations, one for the PRP group and the other for the non-PRP group. It is found that, on average, the predicted JS of workers with PRP schemes is lower than that of those on other pay schemes. However, for high-paid workers, PRP has a positive effect on JS. Interpreting the results, the authors argue that PRP is perceived to be controlling for lower-paid employees, but viewed as supportive rewards by higher-paid workers who derive utility benefit from PRP schemes. The findings suggest that PRP is not universally applicable. While PRP schemes may work effectively for high-paid professions in generating incentives, they can be counterproductive for low-paid occupations.

Questions and Problems

1   What is featured by a censored sample? What is featured by a truncated sample?
2   Why and how do censoring and truncation arise? At what points do censoring and truncation usually come across?
3   Contrast censoring with truncation, and then contrast limited dependent variables associated with censoring and truncation with limited dependent variables associated with discrete choice.
4   Describe and discuss the Tobit model with reference to its implementation and estimation, and the issues in its estimation.
5   What is selection bias? How does selection bias arise? By what means can selection bias be detected and how can selection bias be corrected?
6   What is defined as an inverse Mill's ratio? What is its links to the issues of bias in estimation of censored and truncated samples?
7   Present the Heckman model and illustrate the two-stage estimation procedure, paying attention to the decision process and the level of involvement process.
8   Compare and contrast the Tobit model with the Heckman model with regard to their assumptions and estimation.
9   Collect data on corporate use of derivatives in risk management from various sources, e.g., Thomson ONE Banker and company annual reports. Implement a Tobit model by firstly estimating a probit model for the whole sample for corporate decisions as whether to use derivatives in risk management; and then secondly, if a decision is made for using derivatives, the amount of derivatives used in risk management by the sub-sample of derivatives user firms (using LIMDEP, GAUSS, RATS or other packages). The dependent

variable is corporate use of derivatives. The independent variables may include firm size, financial leverage, the market-to-book ratio, interest coverage ratio, quick ratio, foreign exposure, and an industry dummy.

10 Repeat the above case using Heckman's two-stage procedure. Specifically elaborate on the decision process and the level of involvement process by choosing and justifying the set of independent variables for the two processes respectively.

11 Collect data on outward FDI at the firm level from various sources, e.g., company annual reports and relevant websites and databases. Implement a Tobit model by firstly estimating a probit model for corporate decisions to engage in outward FDI for the whole sample; and then secondly if a decision is made for engaging in FDI overseas, the value of FDI by individual firms in the sub-sample of FDI firms (using LIMDEP, GAUSS, RATS or other packages). The dependent variable is FDI of firms. The independent variables may include firm size, the market-to-book ratio, financial leverage, and an industry dummy.

12 Repeat the above case using Heckman's two-stage procedure. Specifically, elaborate on the process in which a firm decides whether to engage in outward FDI or not and the process for the level of involvement of the firm, by choosing and justifying the set of independent variables for the two processes respectively.

References

Calvo, J.L. (2006), Testing Gibrat's law for small, young and innovating firms, *Small Business Economics*, **26**, 117–123.

Chuang, Y.C. and Lin, C.M. (1999), Foreign direct investment, R&D and spillover efficiency: evidence from Taiwan's manufacturing firms, *Journal of Development Studies*, **35**, 117–137.

Cragg, J.G. (1971), Some statistical models for limited dependent variables with application to the demand for durable goods, *Econometrica*, **39**, 829–844.

Heckman, J.J. (1976), The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models, *Annals of Economic and Social Measurement*, **5**, 475–492.

Heckman, J.J. (1979), Sample selection bias as a specification error, *Econometrica*, **47**, 153–161.

Kneller, R. and Pisu, M. (2007), Industrial linkages and export spillovers from FDI, *World Economy*, **30**, 105–134.

McCausland, W.D., Pouliakas, K. and Theodossiou, I. (2005), Some are punished and some are rewarded: a study of the impact of performance pay on job satisfaction, *International Journal of Manpower*, **26**, 636–659.

Neuman, S. and Oaxaca, R.L. (2005), Wage differentials in the 1990s in Israel: endowments, discrimination, and selectivity, *International Journal of Manpower*, **26**, 217–236.

Power, B. and Reid, G.C. (2005), Flexibility, firm-specific turbulence and the performance of the long-lived small firm, *Review of Industrial Organization*, **26**, 415–443.

Przeworski, A. and Vreeland, J.R. (2000), The effect of IMF programs on economic growth, *Journal of Development Economics*, **62**, 385–421.

Ruffer, R.L. and Holcomb Jr., J.P. (2001), To build or buy: an empirical study of the characteristics affecting a bank's expansion decisions, *Journal of Economics and Business*, **53**, 481–495.

Sinani, E. and Meyer, K.E. (2004), Spillovers of technology transfer from FDI: the case of Estonia, *Journal of Comparative Economics*, **32**, 445–466.

Tobin, J. (1958), Estimation of relationships for limited dependent variables, *Econometrica*, **26**, 24–36.

Villalonga, B. (2004), Does diversification cause the 'diversification discount?' *Financial Management*, **33**, 5–27.

13 Panel data analysis

Panel data covered in this chapter refer to data sets consisting of cross-sectional observations over time, or pooled cross section and time series data. They have two dimensions, one for time and one for the cross-section entity. We have been familiar with the time dimension already, using a subscript t to the variable to stand for the time dimension, with $t = 1, 2, \ldots T$ for T observations at the T time points. For the cross-sectional data, the entity can be individuals, firms, regions or countries, as introduced in the two preceding chapters. The subscript n to the variable is usually adopted to represent the cross-section dimension, with $n = 1, 2, \ldots N$ for N observations for N different entities, e.g., N firms. There can be other forms of panels, as long as there are two dimensions for the observations; but only panel data with a time dimension and a cross-section dimension are covered in this chapter.

Because two dimensions of data are involved, representation and estimation of panel data are different from what we have learned for one-dimension data sets, time series or cross-sectional. Both variations amongst entities and patterns in time are to be examined, which enriches the study. Effects with reference to entity, or time, or both, emerge and have to be dealt with, leading to models with fixed effects, random effects, random coefficients or random parameters, and so on. More information is available in terms of volume, and richer information is available that blends. Statistically, a panel data set can provide more observations to enjoy the large sample status, so the central limit theorem may apply where its respective single dimensional time series or cross-sectional data set fails, making estimation and inference more efficient. The problems associated with the analysis of non-stationary data in time series may be eased from the cross-sectional dimension when the number of independent cross-sectional entities is sufficiently large.

Panel data have been traditionally used in social-economic research. For example, the British Household Panel Survey (BHPS) started its first interviews of households in 1991, with follow-up interviews of these original households being carried out annually. Two of the widely used US panel data sets are Panel

Study of Income Dynamics (PSID) and The National Longitudinal Surveys of Labor Market Experience (NLS). Both were established in the 1960s. There are many other panels, such as The European Community Household Panel (ECHP), The Survey of Income and Program Participation (SIPP) of the US, and The German Socio-Economic Panel (SOEP). Nevertheless, panel data have long been used in financial studies as well as economic research, with panel-like data sets being created intentionally or unintentionally. Regression of one of the firm performance measures, such as market return, on several financial ratios over time constitutes a kind of panel data analysis. A panel data approach to examining purchasing power parity (PPP) has been attempted time and again. One of the British and Irish financial databases, Financial Analysis Made Easy (FAME), contains information on 3.4 million companies in the UK and Ireland, 2.4 million of which are in a detailed format. For the top 2.4 million companies the reports typically include 29 profit and loss account and 63 balance sheet items, cash flow and ratios, credit score and rating, etc. Similar company accounts and financial ratio data are covered by DIANE for France, DAFNE for Germany and Austria, and REACH for the Netherlands, to mention a few; while AMADEUS provides pan-European financial information in the above domains. Thomson ONE Banker extends the coverage of such information to almost all countries in the world. In a sense, the use of panel data in finance and financial economics predated that in social-economic research, and is very extensive too.

This chapter first presents the structure and organisation of panel data sets. Two major features that do not exist with the one dimension time series data or the one dimension cross-sectional data are fixed effects and random effects. So, fixed effects models and random effects models and their estimation are discussed next, followed by random parameter models and their estimation. Then the chapter proceeds to present dynamic panel data models, addresses the issue of bias in parameter estimation and discusses a few approaches to estimating dynamic panel models.

13.1. Structure and organisation of panel data sets

Panel data sets in this chapter are pooled time series and cross-sectional data. We may organise data by stacking individual entities' time series to form a panel with the following structure for an N entity T period and K independent variable panel. In the illustration below, each cross-sectional block is simply a conventional time series arrangement for one entity, with a dependent variable y_t and K independent variables x_{kt}, $k = 1, \ldots K$; $t = 1, \ldots T$. When pooling the time series of the cross-sectional entities together, a panel is formed that requires an additional subscript i to indicate and distinguish the entity. So, for the second entity, for example, the dependent variable is designated y_{2t} and K independent variables are x_{k2t}, $k = 1, \ldots K$; $t = 1, \ldots T$. The above panel is balanced, where all entities possess an observation at all time points. When some entities do not possess an observation at certain time points while some other entities do, the panel

is said to be unbalanced. We focus our study on balanced panel data in this chapter.

i	t	y	x_1	$\ldots$	x_K
1	1	y_{11}	x_{111}	$\ldots$	x_{K11}
1	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
1	T	y_{1T}	x_{11T}	$\ldots$	x_{K1T}
2	1	y_{21}	x_{121}	$\ldots$	x_{K21}
2	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
2	T	y_{2T}	x_{12T}	$\ldots$	x_{K2T}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
N	1	y_{N1}	x_{1N1}	$\ldots$	x_{KN1}
N	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
N	T	y_{NT}	x_{1NT}	$\ldots$	x_{KNT}

A balanced panel data set can be in the form of unstacked data. Although the structure of stacked data by cross-section is straightforward and easy to follow, the structure of unstacked data is usually what we get when downloading data. Therefore, learning the structure of unstacked data is helpful for the transformation and organisation of panel data sets used in empirical estimation and analysis.

t	y_1	x_{11}	$\ldots$	x_{K1}	$\ldots$	y_N	x_{1N}	$\ldots$	x_{KN}
1	y_{11}	x_{111}	$\ldots$	x_{K11}	$\ldots$	y_{N1}	x_{1N1}	$\ldots$	x_{KN1}
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
T	y_{1T}	x_{11T}	$\ldots$	x_{K1T}	$\ldots$	y_{NT}	x_{1NT}	$\ldots$	x_{KNT}

The basic regression equation for panel data is $y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \omega_{it}, i = 1, \ldots N$; $t = 1, \ldots T$. One of the benefits for using panel data is that analysis of panel data can reveal individual variation that is unobservable in time series, and reveal time variation that is unobservable in cross-sections. This implies that the error term ω_{it} may be systematically higher or lower for some individual entities than for other individual entities, or that the error term ω_{it} may be systematically higher or lower for some time periods than for other time periods. The former is accounted for by individual effects and the latter by time effects. Depending on the assumptions followed, the variation in time and cross-sections can be captured by a constant or a random variable, giving rise to fixed effects models and random effects models. Moreover, the coefficient may be different for each individual entity, which corresponds to a random coefficients or random parameters model; and in that case, $\boldsymbol{\beta}$ also has a subscript i to become $\boldsymbol{\beta}_i$. These models and their estimation will be studied in Sections 13.2 and 13.3.

13.2. Fixed effects vs. random effects models

We use the following three equations for the presentation and analysis of panel data models in this section. Dimensions of the matrix or vector are put in the right bottom corner of the matrix or vector, to avoid the usual confusion in presenting panel data. The first is a pooled time series and cross-sections with both the time dimension and the cross-section dimension being explicitly expressed:

$$y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \omega_{it}, \qquad i = 1, \dots N; \quad t = 1, \dots T \tag{13.1}$$

where:

$$\mathbf{X}_{it} = \begin{bmatrix} x_{1it} & \cdots & x_{Kit} \end{bmatrix}_{(K \times 1)} \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \cdots \\ \beta_k \end{bmatrix}_{(1 \times K)}$$

Second, a compact matrix representation for individual entities is as follows:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\omega}_i, \quad i = 1, \dots N \tag{13.2}$$

where:

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ \cdots \\ y_{iT} \end{bmatrix}_{(T \times 1)}, \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{X}_{i1} \\ \cdots \\ \mathbf{X}_{iT} \end{bmatrix} = \begin{bmatrix} x_{1i1} & \cdots & x_{Ki1} \\ \cdots & \cdots & \cdots \\ x_{1iT} & & x_{KiT} \end{bmatrix}_{(T \times K)}$$

$$\text{and} \quad \boldsymbol{\omega}_i = \begin{bmatrix} \omega_{i1} \\ \cdots \\ \omega_{iT} \end{bmatrix}_{(T \times 1)}$$

For any individual entity, $\mathbf{y}_i$ is a $(T \times 1)$ vector for T observations of the dependent variable, $\mathbf{X}_i$ is a $(T \times K)$ matrix of independent variables or regressors with K being the number of independent variables, and $\boldsymbol{\beta}$ is a $(K \times 1)$ vector of coefficients.

Finally, a compact matrix representation for the panel is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\omega} \tag{13.3}$$

where:

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \cdots \\ \mathbf{y}_N \end{bmatrix}_{[(T \times N) \times 1]}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \cdots \\ \mathbf{X}_N \end{bmatrix}_{[(T \times N) \times K]} \quad \text{and} \quad \boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\omega}_1 \\ \cdots \\ \boldsymbol{\omega}_N \end{bmatrix}_{[(T \times N) \times 1]}$$

Various effects associated with the intercept can be formulated by decomposing ω_{it} in different ways. Let us concentrate on individual effects for now. The fixed

effects model assumes that:

$$\omega_{it} = c_i + \varepsilon_{it} \tag{13.4}$$

where c_i is individual-specific and time-invariant unobserved heterogeneity and is a constant for entity i, $\mathrm{Cov}(c_i, \mathbf{X}_{it}) \neq 0$, $\mathrm{Cov}(\varepsilon_{it}, \mathbf{X}_{it}) = 0$, $\mathrm{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$; and ε_{it} is pure residuals uncorrelated with each other and uncorrelated with independent variables. A compact matrix representation for the panel data model with fixed effects is:

$$\mathbf{y} = \mathbf{c} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{13.5}$$

where:

$$\mathbf{c} = \begin{bmatrix} \mathbf{c}_1 \\ \cdots \\ \mathbf{c}_N \end{bmatrix}_{[(T \times N) \times 1]}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \cdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix}_{[(T \times N) \times 1]}, \quad \mathbf{c}_i = \begin{bmatrix} c_i \\ \cdots \\ c_i \end{bmatrix}_{(T \times 1)}$$

$$\text{and} \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \cdots \\ \varepsilon_{iT} \end{bmatrix}_{(T \times 1)}$$

Fixed effects models cannot be readily estimated by the OLS. There are few approaches that augment the OLS, such as resorting to dummies, applying first differencing over time, and performing the within transformation.

To show how the dummy variable approach, the dummy variable least squares (DVLS), works, let $\mathbf{I}_i$ be a $(T \times 1)$ vector of 1's, $\mathbf{I}$ be a $[(T \times N) \times N]$ matrix of $\mathbf{I}_i$ and $\mathbf{0}$, and $\boldsymbol{\delta}$ be a $(N \times 1)$ vector of dummy coefficients, i.e.:

$$\mathbf{I}_i = \begin{bmatrix} 1 \\ \cdots \\ 1 \end{bmatrix}_{(T \times 1)}, \quad \mathbf{I} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I}_N \end{bmatrix}_{[(T \times N) \times N]}, \quad \boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \cdots \\ \delta_N \end{bmatrix}_{(N \times 1)}$$

Then a rearranged representation of equation (13.5) can be estimated by the OLS, which produces unbiased fixed effects estimators:

$$\mathbf{y} = [\mathbf{X}, \mathbf{I}] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} + \boldsymbol{\varepsilon} \tag{13.6}$$

It follows that, $\mathbf{b}$ and $\mathbf{d}$, the DVLS estimators of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, are:

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{d} \end{bmatrix} = \left([\mathbf{X}, \mathbf{I}]'[\mathbf{X}, \mathbf{I}]\right)^{-1}[\mathbf{X}, \mathbf{I}]'\mathbf{y} + \left([\mathbf{X}, \mathbf{I}]'[\mathbf{X}, \mathbf{I}]\right)^{-1}[\mathbf{X}, \mathbf{I}]'\boldsymbol{\varepsilon}$$

$$= \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \end{bmatrix} + \left([\mathbf{X}, \mathbf{I}]'[\mathbf{X}, \mathbf{I}]\right)^{-1}[\mathbf{X}, \mathbf{I}]'\boldsymbol{\varepsilon} \tag{13.7}$$

The fixed effects in a panel data model can be either significant or insignificant. If $\delta_1 = \delta_2 = \cdots \delta_N = \delta$, then δ is simply a common intercept for all the entities within

the panel or for the whole panel. So, restrictions can be imposed on the estimated dummy coefficients. If the restrictions of $\delta_1 = \delta_2 = \cdots \delta_N = \delta$ are rejected, the panel data model is regarded to possess the features of fixed effects and there are fixed effects variations across entities. Otherwise, if the restrictions of $\delta_1 = \delta_2 = \cdots \delta_N = \delta$ are rejected, then there are no fixed effects variations in intercepts and the whole panel has just one common intercept.

When the residual, ε_{it}, is distributed normally, the maximum likelihood procedure can be applied to obtain fixed effects estimators of parameters. The likelihood function and the log likelihood function are as follows:

$$L = \prod_{i=1}^{N} \phi\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta}\right) = \prod_{i=1}^{N} \frac{1}{\sigma_\varepsilon}\phi\left(\frac{\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta}}{\sigma_\varepsilon}\right)$$

$$= \prod_{i=1}^{N} \left(2\pi\sigma_\varepsilon^2\right)^{-T/2}\exp\left(-\left[\frac{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta})}{2\sigma_\varepsilon^2}\right]\right) \quad (13.8)$$

$$LL = \sum_{i=1}^{N} \mathrm{Ln}\left\{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta}}{\sigma_\varepsilon}\right)\right\}$$

$$= \sum_{i=1}^{N} \mathrm{Ln}\left\{\left(2\pi\sigma_\varepsilon^2\right)^{-(N\times T)/2}\exp\left(-\left[\frac{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta})}{2\sigma_\varepsilon^2}\right]\right)\right\}$$

$$= -\frac{N\times T}{2}\mathrm{Ln}\left(2\pi\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{N}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta}\right)'\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{I}_i\boldsymbol{\delta}\right)$$

$$(13.9)$$

where $\phi(z)$ is the density function of normal distributions.

Further, let us consider both individual effects and time effects in the fixed effects panel data model presented as follows:

$$\mathbf{y} = \mathbf{c} + \mathbf{h} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{13.10}$$

where:

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_J \\ \cdots \\ \mathbf{h}_T \end{bmatrix}_{[(T\times N)\times 1]} \quad \text{and} \quad \mathbf{h}_T = \begin{bmatrix} h_1 \\ \cdots \\ h_T \end{bmatrix}_{(T\times 1)}$$

and the rest is the same as in equation (13.5).

The dummy variable approach, the DVLS, can be applied to estimate both individual effects and time effects in the panel data model. However, if we use

one dummy for one entity for all N entities and one dummy for one time period for all T periods, then both individual dummies and time dummies sum to one. So we have to remove one dummy, either time dummy or individual dummy, from estimation. Let us remove one time dummy for the first period, so the first period is set to be the base period. A time dummy coefficient, say h_3 for the third period, indicates that the value of the dependent variable is greater by the extent of h_3 relative to the base period, other things being equal. This dummy variable representation of the panel data model, incorporating both individual effects and time effects, can be expressed by the following equation:

$$\mathbf{y} = [\mathbf{X}, \mathbf{J}] \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \boldsymbol{\zeta} \end{bmatrix} + \boldsymbol{\varepsilon} \tag{13.11}$$

where:

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \ldots & \mathbf{0} & \mathbf{J}_1 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{I}_N & \mathbf{J}_T \end{bmatrix}_{[(T \times N) \times (N \times T)]} , \quad \mathbf{I}_i = \begin{bmatrix} 1 \\ \ldots \\ 1 \end{bmatrix}_{(T \times 1)} ,$$

$$\mathbf{J}_t = \begin{bmatrix} 0 & \ldots & 0 \\ 1 & \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & \ldots & 1 \end{bmatrix}_{[T \times (T-1)]} ,$$

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \ldots \\ \delta_N \end{bmatrix}_{(N \times 1)} \quad \text{and} \quad \boldsymbol{\zeta} = \begin{bmatrix} \zeta_2 \\ \ldots \\ \zeta_T \end{bmatrix}_{[(T-1) \times 1]}$$

Then, $\mathbf{b}$, $\mathbf{d}$ and $\mathbf{z}$, the DVLS estimators of $\boldsymbol{\beta}$, $\boldsymbol{\delta}$ and $\boldsymbol{\zeta}$, can be derived from the following operations:

$$\begin{bmatrix} \mathbf{b} \\ \mathbf{d} \\ \mathbf{z} \end{bmatrix} = ([\mathbf{X}, \mathbf{J}]'[\mathbf{X}, \mathbf{J}])^{-1} [\mathbf{X}, \mathbf{J}]' \mathbf{y} + ([\mathbf{X}, \mathbf{J}]'[\mathbf{X}, \mathbf{J}])^{-1} [\mathbf{X}, \mathbf{J}]' \boldsymbol{\varepsilon}$$

$$= \begin{bmatrix} \boldsymbol{\beta} \\ \boldsymbol{\delta} \\ \boldsymbol{\zeta} \end{bmatrix} + ([\mathbf{X}, \mathbf{J}]'[\mathbf{X}, \mathbf{J}])^{-1} [\mathbf{X}, \mathbf{J}]' \boldsymbol{\varepsilon} \tag{13.12}$$

When the residual, ε_{it}, is assumed to be normally distributed, the maximum likelihood procedure can be applied to obtain fixed effects parameter estimators. The likelihood function and the log likelihood function are described by the

following two representations respectively:

$$L = \phi\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}\right) = \frac{1}{\sigma_\varepsilon}\phi\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}}{\sigma_\varepsilon}\right)$$

$$= (2\pi\sigma_\varepsilon^2)^{-(N\times T)/2}\exp\left(-\left\{\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}\right]'\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}\right]\Big/2\sigma_\varepsilon^2\right\}\right)$$

(13.13)

$$LL = \mathrm{Ln}\left\{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}}{\sigma_\varepsilon}\right)\right\}$$

$$= \mathrm{Ln}\left\{(2\pi\sigma_\varepsilon^2)^{-(N\times T)/2}\exp\left(-\left\{\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}\right]'\right.\right.\right.$$

$$\left.\left.\left.\times\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}\right]\Big/2\sigma_\varepsilon^2\right\}\right)\right\}$$

$$= -\frac{N\times T}{2}\mathrm{Ln}(2\pi\sigma_\varepsilon^2) - \frac{1}{2\sigma_\varepsilon^2}\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}\right]'\left[\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{J}\begin{pmatrix}\delta\\\varsigma\end{pmatrix}\right] \quad (13.14)$$

The within transformation approach can also produce unbiased OLS estimators. Let us consider a fixed effects model with individual effects only:

$$y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + c_i + \varepsilon_{it}, \quad i = 1, \dots N; \quad t = 1, \dots T \tag{13.15}$$

The within transformation is to average the variables over time for each entity, and then subtract the entity average from their counterparts in the original equation at all time points, which leads to the cancelling out of the individual effects:

$$y_{it} - \bar{y}_i = (\mathbf{X}_{it} - \overline{\mathbf{X}}_i)\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i, \quad i = 1, \dots N; \quad t = 1, \dots T \tag{13.16}$$

The individual effects can also be removed by applying first differences over time in equation (13.15):

$$\Delta y_{it} = \Delta\mathbf{X}_{it}\boldsymbol{\beta} + \Delta\varepsilon_{it}, \quad i = 1, \dots N; \quad t = 1, \dots T \tag{13.17}$$

Since the individual effects have been removed, there is no concern on the correlation between the individual effects and the independent variables which causes bias in estimation. The OLS estimator is unbiased, consistent but inefficient since the off- diagonal elements of the covariance matrix are no longer zero, which is not taken into account by the OLS. Remedies and improvement measures can

be too complicated to attempt. Therefore, applications of other readily available procedures are advised.

The random effects model with individual effects assumes that:

$$\omega_{it} = \mu_i + \varepsilon_{it} \tag{13.18}$$

where μ_i is a random variable, $E(\mu_i) = 0$, $\text{Var}(\mu_i) = \sigma_\mu^2$, $\text{Cov}(\mu_i, \mathbf{X}_{it}) = 0$, $\text{Cov}(\mu_i, \varepsilon_{it}) = 0$, $E(\varepsilon_{it}) = 0$, $\text{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$; and ε_{it} is pure residuals uncorrelated with each other and uncorrelated with independent variables. A compact matrix representation for the panel data model with random individual effects is:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\varepsilon} \tag{13.19}$$

where:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \dots \\ \boldsymbol{\mu}_N \end{bmatrix}_{[(T \times N) \times 1]} \quad , \quad \boldsymbol{\mu}_i = \begin{bmatrix} \mu_1 \\ \dots \\ \mu_N \end{bmatrix}_{(T \times 1)} \quad , \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \dots \\ \boldsymbol{\varepsilon}_N \end{bmatrix}_{[(T \times N) \times 1]}$$

$$\text{and} \quad \boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i1} \\ \dots \\ \varepsilon_{iT} \end{bmatrix}_{(T \times 1)}$$

Taking into consideration the non-zero within entity covariance or non-zero off-diagonal elements of the entity covariance matrix, and the heteroskadasticity arising from the heterogeneity of effects, its random effects estimators can be derived by the GLS as follows:

$$\mathbf{b} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} + \boldsymbol{\varepsilon} \right)$$

$$= \boldsymbol{\beta} + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X} \right)^{-1} \mathbf{X}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} + \boldsymbol{\varepsilon} \right) \tag{13.20}$$

with the covariance matrix for the panel being:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & & & \\ & \boldsymbol{\Sigma}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\Sigma}_N \end{bmatrix}_{[(T \times N) \times (T \times N)]} \tag{13.21}$$

and the within entity covariance matrix being:

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_\mu^2 + \sigma_\varepsilon^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_\varepsilon^2 & \dots & \sigma_\mu^2 \\ \dots & \dots & \dots & \dots \\ \sigma_\mu^2 & \sigma_\mu^2 & \dots & \sigma_\mu^2 + \sigma_\varepsilon^2 \end{bmatrix}_{(T \times T)} \tag{13.22}$$

As having been pointed out in Chapter 3 that the above covariance matrices are not readily available and have to be estimated first. The customary means is

to apply the OLS first for the purpose of calculating the residuals, from which the covariance matrix can be derived. It is a rather tedious, though not difficult, procedure. Therefore, other methods, such as the maximum likelihood procedure, are usually employed in practice.

When normal distributions are assumed for the residual ε_{it}, the maximum likelihood procedure can be applied to obtain random effects estimators of parameters. The corresponding likelihood function and the log likelihood function are presented as follows:

$$L = (2\pi)^{-(N \times T)/2} \left(|\mathbf{\Sigma}^{-1}| \right)^{1/2} \exp\left(-\left[(\mathbf{y} - \mathbf{X\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{X\beta}) / 2 \right] \right)$$

$$= \prod_{i=1}^{N} (2\pi)^{-T/2} \left(|\mathbf{\Sigma}_i^{-1}| \right)^{1/2} \exp\left(-\left[(\mathbf{y}_i - \mathbf{X}_i\mathbf{\beta})' \mathbf{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\mathbf{\beta}) / 2 \right] \right)$$

(13.23)

$$LL = \mathrm{Ln} \left\{ (2\pi)^{(N \times T)/2} \left(|\mathbf{\Sigma}^{-1}| \right)^{1/2} \exp\left(-\left[(\mathbf{y} - \mathbf{X\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{X\beta}) / 2 \right] \right) \right\}$$

$$= \frac{1}{2} \left[-N \times T \times \mathrm{Ln}(2\pi) + \mathrm{Ln}\left(|\mathbf{\Sigma}^{-1}| \right) - (\mathbf{y} - \mathbf{X\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{X\beta}) \right]$$

$$= \frac{1}{2} \sum_{i=1}^{N} \left[-T \times \mathrm{Ln}(2\pi) + \mathrm{Ln}\left(|\mathbf{\Sigma}_i^{-1}| \right) - (\mathbf{y}_i - \mathbf{X}_i\mathbf{\beta})' \mathbf{\Sigma}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i\mathbf{\beta}) \right]$$

(13.24)

The random effects model with both individual effects and time effects assumes that:

$$\omega_{it} = \mu_i + \tau_t + \varepsilon_{it}$$

(13.25)

where μ_i is a random variable with $E(\mu_i) = 0$, $\mathrm{Var}(\mu_i) = \sigma_\mu^2$, $\mathrm{Cov}(\mu_i, \mathbf{X}_{it}) = 0$, $\mathrm{Cov}(\mu_i, \varepsilon_{it}) = 0$; τ_t is a random variable with $E(\tau_t) = 0$, $\mathrm{Var}(\tau_t) = \sigma_\mu^2$, $\mathrm{Cov}(\tau_t, \mathbf{X}_{it}) = 0$, $\mathrm{Cov}(\tau_t, \varepsilon_{it}) = 0$, $E(\varepsilon_{it}) = 0$, $\mathrm{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$; and ε_{it} is pure residuals uncorrelated with each other and uncorrelated with independent variables. A compact matrix representation for the panel data model with both random individual effects and random time effects is:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{\mu} + \mathbf{\tau} + \mathbf{\varepsilon}$$

(13.26)

where:

$$\mathbf{\tau} = \begin{bmatrix} \mathbf{\tau}_t \\ \cdots \\ \mathbf{\tau}_t \end{bmatrix}_{[(T \times N) \times 1]} \quad \text{and} \quad \mathbf{\tau}_t = \begin{bmatrix} \tau_1 \\ \cdots \\ \tau_T \end{bmatrix}_{(T \times 1)}$$

and the rest is the same as in equation (13.19).

Taking into consideration the non-zero within entity covariance or non-zero off-diagonal elements of the entity covariance matrix, and the heteroskadasticity arising from the heterogeneity of effects, its random effects estimators of $\boldsymbol{\beta}$, $\mathbf{b}$, can be derived by the GLS as follows:

$$\mathbf{b} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}+\boldsymbol{\tau}+\boldsymbol{\varepsilon})$$
$$= \boldsymbol{\beta} + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}+\boldsymbol{\tau}+\boldsymbol{\varepsilon}) \tag{13.27}$$

with the covariance matrix for the panel being:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} & \cdots & \boldsymbol{\Sigma}_{1N} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 & & \\ \cdots & & \ddots & \\ \boldsymbol{\Sigma}_{N1} & & & \boldsymbol{\Sigma}_N \end{bmatrix}_{[(T\times N)\times(T\times N)]} \tag{13.28}$$

where the within entity covariance matrix is:

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \sigma_\mu^2 + \sigma_\varepsilon^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 + \sigma_\varepsilon^2 & \cdots & \sigma_\mu^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 + \sigma_\varepsilon^2 \end{bmatrix}_{(T\times T)} \tag{13.29}$$

and the between entities covariance matrix is:

$$\boldsymbol{\Sigma}_{ij} = \begin{bmatrix} \sigma_\tau^2 & 0 & \cdots & 0 \\ 0 & \sigma_\tau^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & & \cdots & \sigma_\tau^2 \end{bmatrix}_{(T\times T)} \tag{13.30}$$

With normally distributed residuals ε_{it}, the maximum likelihood procedure can be applied. The corresponding likelihood function and the log likelihood function are as follows:

$$L = (2\pi)^{-(N\times T)/2}\left(\left|\boldsymbol{\Sigma}^{-1}\right|\right)^{1/2}\exp\left(-\left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2\right]\right) \tag{13.31}$$

$$LL = \mathrm{Ln}\left\{(2\pi)^{(N\times T)/2}\left(\left|\boldsymbol{\Sigma}^{-1}\right|\right)^{1/2}\exp\left(-\left[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})/2\right]\right)\right\}$$
$$= -\frac{1}{2}\left[-N\times T\times\mathrm{Ln}(2\pi)+\mathrm{Ln}\left(\left|\boldsymbol{\Sigma}^{-1}\right|\right)-(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\right] \tag{13.32}$$

Both fixed effects and random effects models can be estimated by the maximum likelihood method. When residuals obey normal distributions, the maximum likelihood method produces the same estimated parameters as the OLS.

13.3. Random parameter models

The random parameter or random coefficient model with individual heterogeneity assumes:

$$\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\mu}_i \tag{13.33}$$

in the panel equation of:

$$y_{it} = \mathbf{X}_{it}(\boldsymbol{\beta} + \boldsymbol{\mu}_i) + \varepsilon_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{X}_{it}\boldsymbol{\mu}_i + \varepsilon_{it}, \quad i = 1,\dots N; \quad t = 1,\dots T \tag{13.34}$$

where:

$$\boldsymbol{\mu}_i = \begin{bmatrix} \mu_{i1} \\ \dots \\ \mu_{iK} \end{bmatrix}_{(K \times 1)}$$

So, a compact matrix representation of the random parameter model for individual entities takes the following form:

$$\mathbf{y}_i = \mathbf{X}_i(\boldsymbol{\beta} + \boldsymbol{\mu}_i) + \boldsymbol{\varepsilon}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i, \quad i = 1,\dots N \tag{13.35}$$

Let us construct a new matrix of independent variables $\mathbf{Z}$ that stack in such a fashion:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{X}_N \end{bmatrix}_{[(T \times N) \times (K \times N)]} \tag{13.36}$$

Then, a compact matrix representation for the panel data model with individual heterogeneity in random parameters can be expressed as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu} + \boldsymbol{\varepsilon} \tag{13.37}$$

where:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \dots \\ \boldsymbol{\mu}_N \end{bmatrix}_{[(K \times N) \times 1]}$$

with $E(\mu_i) = 0$, $\text{Var}(\mu_i) = \sigma_\mu^2$, $\text{Cov}(\mu_i, \mathbf{X}_{it}) = 0$, $\text{Cov}(\mu_i, \varepsilon_{it}) = 0$, $E(\varepsilon_{it}) = 0$, $\text{Var}(\varepsilon_{it}) = \sigma_\varepsilon^2$; and ε_{it} is pure residuals uncorrelated with each other and uncorrelated with independent variables.

Taking into consideration the non-zero within entity covariance or non-zero off-diagonal elements of the entity covariance matrix, and the heteroskadasticity

arising from the heterogeneity of random parameters, the random parameter estimators of $\boldsymbol{\beta}$, $\mathbf{b}$, can be derived by the GLS as follows:

$$\mathbf{b} = \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z}\boldsymbol{\mu}) + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}$$

$$= \boldsymbol{\beta} + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\mathbf{Z}\boldsymbol{\mu}) + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} \qquad (13.38)$$

with the covariance matrix for the panel being:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & & & \\ & \boldsymbol{\Sigma}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{\Sigma}_N \end{bmatrix}_{[(T \times N) \times (T \times N)]} \qquad (13.39)$$

and the within entity covariance matrix being:

$$\boldsymbol{\Sigma}_i = \sigma_\varepsilon^2 \mathbf{I} + \mathbf{X}_i \boldsymbol{\Sigma}_\mu \mathbf{X}_i' \qquad (13.40)$$

where:

$$\boldsymbol{\Sigma}_\mu = \begin{bmatrix} \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \end{bmatrix}_{(T \times T)} \qquad (13.41)$$

The unconditional likelihood function and log likelihood function of this model, assuming normal distributions for the residual, are:

$$L = (2\pi)^{-(N \times T)/2} \left(|\boldsymbol{\Sigma}^{-1}|\right)^{1/2} \exp\left(-\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/2\right]\right)$$

$$= \prod_{i=1}^{N} (2\pi)^{-T/2} \left(|\boldsymbol{\Sigma}_i^{-1}|\right)^{1/2} \exp\left(-\left[(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})/2\right]\right)$$

$$(13.42)$$

$$LL = \mathrm{Ln}\left\{(2\pi)^{-(N \times T)/2} \left(|\boldsymbol{\Sigma}^{-1}|\right)^{1/2} \exp\left(-\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})/2\right]\right)\right\}$$

$$= \frac{1}{2}\left[-N \times T \times \mathrm{Ln}(2\pi) + \mathrm{Ln}\left(|\boldsymbol{\Sigma}^{-1}|\right) - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right]$$

$$= \frac{1}{2}\sum_{i=1}^{N}\left[-T \times \mathrm{Ln}(2\pi) + \mathrm{Ln}\left(|\boldsymbol{\Sigma}_i^{-1}|\right) - (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})\right]$$

$$(13.43)$$

respectively. Its conditional likelihood function and log likelihood function are:

$$L = \prod_{i=1}^{N} \phi\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}\right) = \prod_{i=1}^{N} \frac{1}{\sigma_\varepsilon}\phi\left(\frac{\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right)$$

$$= \prod_{i=1}^{N} \left(2\pi\sigma_\varepsilon^2\right)^{-T/2} \exp\left(-\left[\frac{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2\sigma_\varepsilon^2}\right]\right) \quad (13.44)$$

$$LL = \sum_{i=1}^{N} \mathrm{Ln}\left\{\frac{1}{\sigma_\varepsilon}\phi\left(\frac{\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}}{\sigma_\varepsilon}\right)\right\}$$

$$= \sum_{i=1}^{N} \mathrm{Ln}\left\{\left(2\pi\sigma_\varepsilon^2\right)^{-T/2} \exp\left(-\left[\frac{(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})}{2\sigma_\varepsilon^2}\right]\right)\right\}$$

$$= -\frac{N \times T}{2}\mathrm{Ln}\left(2\pi\sigma_\varepsilon^2\right) - \frac{1}{2\sigma_\varepsilon^2}\sum_{i=1}^{N}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) \quad (13.45)$$

respectively.

Random parameter models with regard to both individual and time entities take the following form:

$$y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + \mathbf{X}_{it}\boldsymbol{\mu}_i + \mathbf{X}_{it}\boldsymbol{\tau}_t + \varepsilon_{it}, \qquad i = 1,\dots N; \qquad t = 1,\dots T \quad (13.46)$$

where:

$$\boldsymbol{\tau}_t = \begin{bmatrix} \tau_{t1} \\ \dots \\ \tau_{tK} \end{bmatrix}_{(K \times 1)}$$

We may construct a new matrix of independent variables $\mathbf{W}_i$ for entity i and a new matrix of independent variables for the panel as follows:

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{X}_{i1} & \mathbf{0} & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{X}_{iT} \end{bmatrix}_{[T \times (T \times K)]} \quad (13.47)$$

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \dots \\ \mathbf{W}_N \end{bmatrix}_{[(T \times N) \times (T \times K)]} \quad (13.48)$$

A compact matrix representation for the panel data model with both individual and time heterogeneity in random parameters becomes:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\mu} + \mathbf{W}\boldsymbol{\tau} + \boldsymbol{\varepsilon} \quad (13.49)$$

where:

$$\boldsymbol{\tau} = \begin{bmatrix} \boldsymbol{\tau}_1 \\ \dots \\ \boldsymbol{\tau}_T \end{bmatrix}_{[(K \times T) \times 1]}$$

The GLS estimators of $\boldsymbol{\beta}$, $\mathbf{b}$, can be obtained as follows:

$$\begin{aligned}
\mathbf{b} &= \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{y} + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{Z}\boldsymbol{\mu}\right) \\
&\quad + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{W}\boldsymbol{\tau}\right) + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta} + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{Z}\boldsymbol{\mu}\right) + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\left(\mathbf{W}\boldsymbol{\tau}\right) \\
&\quad + \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}
\end{aligned} \tag{13.50}$$

with the covariance matrix for the panel being:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{\Sigma}_{12} & \dots & \boldsymbol{\Sigma}_{1N} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_2 & & \\ \dots & & \ddots & \\ \boldsymbol{\Sigma}_{N1} & & & \boldsymbol{\Sigma}_N \end{bmatrix}_{[(T \times N) \times (T \times N)]} \tag{13.51}$$

where the within entity covariance matrix is:

$$\boldsymbol{\Sigma}_i = \sigma_{\varepsilon}^2\mathbf{I} + \mathbf{X}_i\boldsymbol{\Sigma}_{\mu}\mathbf{X}_i' \tag{13.52}$$

with:

$$\boldsymbol{\Sigma}_{\mu} = \begin{bmatrix} \sigma_{\mu}^2 & \sigma_{\mu}^2 & \dots & \sigma_{\mu}^2 \\ \sigma_{\mu}^2 & \sigma_{\mu}^2 & \dots & \sigma_{\mu}^2 \\ \dots & \dots & \dots & \dots \\ \sigma_{\mu}^2 & \sigma_{\mu}^2 & \dots & \sigma_{\mu}^2 \end{bmatrix}_{(T \times T)} \tag{13.53}$$

and the between entities covariance matrix is:

$$\boldsymbol{\Sigma}_{ij} = \begin{bmatrix} \sigma_{\tau}^2 & 0 & \dots & 0 \\ 0 & \sigma_{\tau}^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & & \dots & \sigma_{\tau}^2 \end{bmatrix}_{(N \times N)} \tag{13.54}$$

Further assuming normal distributions for the residual, the unconditional likelihood function and log likelihood function of this model are obtained as follows:

$$L = (2\pi)^{(N \times T)/2} \left(|\mathbf{\Sigma}^{-1}| \right)^{1/2} \exp\left(-\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) / 2 \right] \right) \quad (13.55)$$

$$LL = \mathrm{Ln}\left\{ (2\pi)^{-(N \times T)/2} \left(|\mathbf{\Sigma}^{-1}| \right)^{1/2} \exp\left(-\left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \Big/ 2 \right] \right) \right\}$$

$$= \frac{1}{2}\left[-N \times T \times \mathrm{Ln}(2\pi) + \mathrm{Ln}\left(|\mathbf{\Sigma}^{-1}| \right) - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{\Sigma}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right]$$
$$(13.56)$$

respectively. Its conditional likelihood function and log likelihood function are:

$$L = \prod_{i=1}^{N} \prod_{t=1}^{T} \phi \left(y_{it} - \mathbf{X}_{it} \boldsymbol{\beta}_{it} \right) = \prod_{i=1}^{N} \prod_{t=1}^{T} \frac{1}{\sigma_\varepsilon} \phi \left(\frac{y_{it} - \mathbf{X}_{it} \boldsymbol{\beta}_{it}}{\sigma_\varepsilon} \right)$$

$$= \prod_{i=1}^{N} \prod_{t=1}^{T} \left(2\pi \sigma_\varepsilon^2 \right)^{-1/2} \exp\left(-\left[\frac{(y_{it} - \mathbf{X}_{it}\boldsymbol{\beta}_{it})'(y_{it} - \mathbf{X}_{it}\boldsymbol{\beta}_{it})}{2\sigma_\varepsilon^2} \right] \right) \quad (13.57)$$

$$LL = -\frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\mathrm{Ln}\left(2\pi \sigma_\varepsilon^2 \right) + \frac{(y_{it} - \mathbf{X}_{it}\boldsymbol{\beta}_{it})'(y_{it} - \mathbf{X}_{it}\boldsymbol{\beta}_{it})}{\sigma_\varepsilon^2} \right] \quad (13.58)$$

respectively.

13.4. Dynamic panel data analysis

Virtually all time series can be extended to cross-sections and all cross-section data can be pooled over time to form a panel. Therefore, those features in time series and models for time series analysis, such as autoregressive processes, unit roots and cointegration, would appear in panels and panel data analysis. Similarly, those features in cross-sectional data and their modelling, such as binary choice and discrete choice models and analysis of truncated and censored data, would also appear in panels and panel data analysis. While the use of panel data enjoys certain advantages over the use of a time series or a cross-sectional data set on the one hand, it may complicate the modelling and analysis on the other hand. For example, when the dependent variable follows a simple AR(1) process, the simplest of dynamic models, the standard within transformation estimator with fixed individual effects is biased and, consequently, alternative estimation procedures need to be developed and applied under pertinent circumstances and with appropriate assumptions. This section hence briefly introduces several generally used procedures for dynamic panel data, focusing on the specific issues in estimation with regard to biasedness and efficiency, and the measures to deal with these issues.

A dynamic panel data model is a model in which the lagged dependent variable appears on the right-hand side of the equation. Nickell (1981) has addressed the issue of biases in dynamic panel models. Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998) put forward several GMM procedures, which have promoted the application of dynamic panel data models in empirical research in recent years. The simplest dynamic panel model is where the dependent variable follows an AR(1) process. The following presentation augments equation (13.1) by including the lagged dependent variable, $y_{i,t-1}$, along with $\mathbf{X}_{it}$, as a regressor:

$$y_{it} = \rho y_{i,t-1} + \mathbf{X}_{it}\boldsymbol{\beta} + \omega_{it} \quad i = 1, \ldots N; \quad t = 1, \ldots T \tag{13.59}$$

where $|\rho| < 1$ to ensure stationarity.

For simplicity and without affecting the outcome, we do not include exogenous independent variables $\mathbf{X}_{it}$ for the moment and only consider individual effects in the discussion. Hence equation (3.59) becomes:

$$y_{it} = \rho y_{i,t-1} + \mu_i + \varepsilon_{it}, \quad i = 1, \ldots N; \quad t = 1, \ldots T \tag{13.60}$$

We do not specify whether μ_i is fixed effects or random effects; but we will see that it is irrelevant and the effects can always be regarded as fixed. Nickell (1981) shows that the LSDV estimator of equation (13.60) is biased. It is because the correlation between the lagged dependent variable and the transformed residual does not disappear. After the within transformation, equation (13.60) becomes:

$$y_{it} - \bar{y}_i = \rho\left(y_{i,t-1} - \bar{y}_{i,-1}\right) + \varepsilon_{it} - \bar{\varepsilon}_i, \quad i = 1, \ldots N; \quad t = 1, \ldots T \tag{13.61}$$

It is clear that $\bar{\varepsilon}_i$ is correlated with $y_{i,t-1}$, so $\hat{\rho}$, the LSDV estimator of ρ, is biased, since:

$$\rho = \frac{\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{it} - \bar{y}_i)(y_{i,t-1} - \bar{y}_{i,-1})}{\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{i,t-1} - \bar{y}_{i,-1})^2} + \frac{\sum_{t=1}^{T}\sum_{i=1}^{N}(\varepsilon_{it} - \bar{\varepsilon}_i)(y_{i,t-1} - \bar{y}_{i,-1})}{\sum_{t=1}^{T}\sum_{i=1}^{N}(y_{i,t-1} - \bar{y}_{i,-1})^2} \tag{13.62}$$

The last term does not converge to zero when $N \to \infty$ but T is not large. Nickell (1981) demonstrates that:

$$p\lim_{N\to\infty}(\hat{\rho} - \rho) = \left\{\frac{2\rho}{1-\rho^2} - \left[\frac{1+\rho}{T-1}\left(1 - \frac{1}{T}\frac{\left(1-\rho^T\right)}{(1-\rho)}\right)\right]^{-1}\right\}^{-1} \tag{13.63}$$

For a reasonably large T, the above is approximately:

$$p\lim_{N\to\infty}(\hat{\rho} - \rho) \approx \frac{1+\rho}{1-T} \tag{13.64}$$

which means the estimator is almost unbiased. However, for small T the bias can be serious. For example:

$$p \lim_{N \to \infty} (\hat{\rho} - \rho) = -\frac{1+\rho}{2}, \quad T = 2 \tag{13.65}$$

To solve this problem, a number of GMM procedures have been proposed, amongst which are the first differenced GMM developed and discussed in Arellano and Bond (1991), Areliano and Bover (1995) and Blundell and Bond (1998), and the system GMM proposed by Blundell and Bond (1998). Let us consider the first differenced GMM first. Taking the difference operation once on all the variables in equation (13.60) yields:

$$\Delta y_{it} = \rho \Delta y_{i,t-1} + \Delta \varepsilon_{it}, \quad i = 1, \dots N; \quad t = 2, \dots T \tag{13.66}$$

where $\Delta y_{it} = y_{it} - y_{i,t-1}$ and $\Delta \varepsilon_{it} = \varepsilon_{it} - \varepsilon_{i,t-1}$.

The idea of the first differenced GMM is as follows. Although $\Delta \varepsilon_{it}$ is correlated with $\Delta y_{i,t-1}$, it is not correlated with $y_{i,t-2}$ or the dependent variable at any more lags, i.e., $y_{i,t-j}$ for $j \geq 2$. That is:

$$E\left(\Delta \varepsilon_{it} y_{i,t-j}\right) = 0, \quad j = 2, \dots t-1, \quad t = 3, \dots T \tag{13.67}$$

Therefore, $y_{i,t-j}$ for $j \geq 2$ and any of the linear combinations can serve as instruments in GMM procedures. Let us define the following matrices and vectors:

$$\mathbf{Z}_i = \begin{bmatrix} y_{i,1} & & & & \\ & y_{i,1} & y_{i,2} & & \\ & & & \ddots & \\ & & & & y_{i,1} & y_{i,2} & \cdots & y_{i,T-2} \end{bmatrix}_{\left[(T-2) \times \frac{(T-2)(T-1)}{2}\right]} \tag{13.68}$$

$$\boldsymbol{\varepsilon}_i = \begin{bmatrix} \varepsilon_{i,3} \\ \cdots \\ \varepsilon_{i,T} \end{bmatrix}_{[(T-2) \times 1]}, \quad \mathbf{y}_i = \begin{bmatrix} y_{i,3} \\ \cdots \\ y_{i,T} \end{bmatrix}_{[(T-2) \times 1]}, \quad \text{and} \quad \mathbf{y}_{i,-1} = \begin{bmatrix} y_{i,2} \\ \cdots \\ y_{i,T-1} \end{bmatrix}_{[(T-2) \times 1]}$$

Then equation (13.67) can be represented by the following moment conditions:

$$E\left(\mathbf{Z}_i' \Delta \boldsymbol{\varepsilon}_i\right) = 0 \tag{13.69}$$

The GMM estimator of the model is derived from minimising the following:

$$\left(\sum_{i=1}^N \mathbf{Z}_i' \Delta \boldsymbol{\varepsilon}_i\right)' \mathbf{W}_N \left(\sum_{i=1}^N \mathbf{Z}_i' \Delta \boldsymbol{\varepsilon}_i\right) = \left(\sum_{i=1}^N \Delta \boldsymbol{\varepsilon}_i' \mathbf{Z}_i\right) \mathbf{W}_N \left(\sum_{i=1}^N \mathbf{Z}_i' \Delta \boldsymbol{\varepsilon}_i\right) \tag{13.70}$$

where $\mathbf{W}_N$ is a weight matrix.

Blundell and Bond (1998) point out that alternative choices for the weight $\mathbf{W}_N$ give rise to a set of GMM estimators, all of which are consistent for large N and finite T, but which differ in their asymptotic efficiency. One of the optimal weight matrices is:

$$\mathbf{W}_N = \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}'_i \Delta \hat{\boldsymbol{\varepsilon}}_i \mathbf{Z}'_i \Delta \hat{\boldsymbol{\varepsilon}}_i \right)^{-1} \tag{13.71}$$

where $\Delta \hat{\boldsymbol{\varepsilon}}_i$ is the residual obtained from an initial consistent estimator. Using the weight matrix $\mathbf{W}_N$ of equation (13.71), the first differenced GMM estimator is derived as:

$$\hat{\rho}_{diff} = \left[\left(\sum_{i=1}^{N} \Delta \mathbf{y}'_{i,-1} \mathbf{Z}_i \right) \mathbf{W}_N \left(\sum_{i=1}^{N} \mathbf{Z}'_i \Delta \mathbf{y}_{i,-1} \right) \right]^{-1}$$
$$\times \left[\left(\sum_{i=1}^{N} \Delta \mathbf{y}'_{i,-1} \mathbf{Z}_i \right) \mathbf{W}_N \left(\sum_{i=1}^{N} \mathbf{Z}'_i \Delta \mathbf{y}_i \right) \right] \tag{13.72}$$

When exogenous independent variables are involved as in the model represented by equation (13.61), the first differenced GMM estimator vector can be easily derived, analogue to equation (13.69). Let us define:

$$\boldsymbol{\beta}^e = \begin{bmatrix} \rho \\ \beta_1 \\ \dots \\ \beta_k \end{bmatrix}_{[1 \times (K+1)]} , \quad \mathbf{x}^e_{i,-1} = \begin{bmatrix} y_{i,2} & x_{1,i,1} & \cdots & x_{k,i,1} \\ \dots & & & \\ y_{i,T-1} & x_{1,i,T-2} & & x_{k,i,T-2} \end{bmatrix}_{[(T-2) \times (K+1)]} \tag{13.73}$$

The first differenced GMM estimator vector is then derived as:

$$\hat{\boldsymbol{\beta}}^e_{\text{diff}} = \left[\left(\sum_{i=1}^{N} \left(\Delta \mathbf{x}^e_{i,-1} \right)' \mathbf{Z}_i \right) \mathbf{W}_N \left(\sum_{i=1}^{N} \mathbf{Z}'_i \Delta \mathbf{x}^e_{i,-1} \right) \right]^{-1}$$
$$\times \left[\left(\sum_{i=1}^{N} \left(\Delta \mathbf{x}^e_{i,-1} \right)' \mathbf{Z}_i \right) \mathbf{W}_N \left(\sum_{i=1}^{N} \mathbf{Z}'_i \Delta \mathbf{y}_i \right) \right] \tag{13.74}$$

To test for the validity of the procedure, the Sargan test for over-identifying has been developed as follows:

$$S = N \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}'_i \Delta \varepsilon_i \right)' \mathbf{W}_N \left(\frac{1}{N} \sum_{i=1}^{N} \mathbf{Z}'_i \Delta \varepsilon_i \right) \tag{13.75}$$

The null hypothesis for this test is that the instruments are not correlated with the errors in the first-differenced equation. The test statistic obeys a χ^2-distribution

with m degrees of freedom under the null, where m is equal to the number of instruments subtracting the number of parameters in the model.

In addition to lagged levels of the dependent variable as instruments for equations in first differences, Blundell and Bond (1998) propose the use of lagged first differences of the dependent variable as instruments for equations in levels. This is referred to as the system GMM. Imposed are the following additional moment restrictions:

$$E\left(\omega_{it}\Delta y_{i,t-1}\right) = 0, \quad t = 4, \dots T \tag{13.76}$$

$$E\left(\omega_{i3}\Delta y_{i,2}\right) = 0 \tag{13.77}$$

where $\omega_{it} = \mu_i + \varepsilon_{it}$. Accordingly, the instrument matrix using the conditions expressed in equation (13.76) and equation (13.77) is as follows:

$$\mathbf{Z}_i^+ = \begin{bmatrix} \mathbf{Z}_i & 0 & 0 & \dots & 0 \\ 0 & \Delta y_{i2} & & & \dots \\ \dots & & \Delta y_{i3} & & \\ & & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & \Delta y_{i,T-1} \end{bmatrix}_{\left\{[2(T-2)]\times\left[\frac{(T-2)(T+1)}{2}\right]\right\}} \tag{13.78}$$

with $\mathbf{Z}_i$ being defined by equation (13.67). Analogue to the first differenced GMM procedure, the system GMM estimator is derived from minimising the following function:

$$\left(\sum_{i=1}^{N}(\mathbf{Z}_i^+)'\Delta\boldsymbol{\varepsilon}_i\right)'\mathbf{W}_N\left(\sum_{i=1}^{N}(\mathbf{Z}_i^+)'\Delta\boldsymbol{\varepsilon}_i\right) = \left(\sum_{i=1}^{N}\Delta\boldsymbol{\varepsilon}_i'\mathbf{Z}_i^+\right)\mathbf{W}_N\left(\sum_{i=1}^{N}(\mathbf{Z}_i^+)'\Delta\boldsymbol{\varepsilon}_i\right) \tag{13.79}$$

The system GMM estimator is then given by:

$$\hat{\rho}_{sys} = \left[\left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{Z}_i^+\right)\mathbf{W}_N\left(\sum_{i=1}^{N}(\mathbf{Z}_i^+)'\mathbf{y}_{i,-1}\right)\right]^{-1}$$
$$\times\left[\left(\sum_{i=1}^{N}\mathbf{y}_{i,-1}'\mathbf{Z}_i^+\right)\mathbf{W}_N\left(\sum_{i=1}^{N}(\mathbf{Z}_i^+)'\mathbf{y}_i\right)\right] \tag{13.80}$$

Similar to the first differenced GMM case, when exogenous independent variables are involved, the system GMM estimator vector can be derived as follows:

$$\hat{\boldsymbol{\beta}}_{sys}^{e} = \left[\left(\sum_{i=1}^{N}(\mathbf{x}_{i,-1}^{e})'\mathbf{Z}_i^+\right)\mathbf{W}_N\left(\sum_{i=1}^{N}(\mathbf{Z}_i^+)'\mathbf{x}_{i,-1}^{e}\right)\right]^{-1}$$
$$\times\left[\left(\sum_{i=1}^{N}(\mathbf{x}_{i,-1}^{e})'\mathbf{Z}_i^+\right)\mathbf{W}_N\left(\sum_{i=1}^{N}(\mathbf{Z}_i^+)'\mathbf{y}_i\right)\right] \tag{13.81}$$

13.5. Examples and cases

Example 13.1

Investment is one of the most important corporate activities taken by firms and pursued by CEOs on behalf of the firms and shareholders. Personal characteristics of CEOs may therefore influence corporate investment behaviour. Malmendier and Tate (2005) argue that managerial overconfidence can account for corporate investment distortions and hypothesise that overconfident managers overestimate the returns to their investment projects and view external funds as unduly costly. Using panel data on personal portfolio and corporate investment decisions of Forbes 500 CEOs, they test the overconfidence hypothesis. The two well-publicised traditional explanations for investment distortions are the misalignment of managerial and shareholders interests and asymmetric information between corporate insiders and the capital market. Both cause investment to be sensitive to the level of cash flows in the firm. The alternative explanation proposed by Malmendier and Tate (2005) in this study relates corporate investment decisions to personal characteristics of the CEO of the firm. They argue that overconfident CEOs systematically overestimate the return to their investment projects. Consequently, they overinvest if they have sufficient internal funds for investment and are not disciplined by the capital market or corporate governance mechanisms. They curb their investment if they do not have sufficient internal funds, since they are reluctant to issue new equity which undervalues the stock of their company by the market and new investors.

   Measures of CEO overconfidence in this study are constructed based on the overexposure of CEOs to the idiosyncratic risk of their firms. The first two measures, Holder 67 and Longholder, are linked to the timing of options' exercises. Risk-averse CEOs should exercise their options early given a sufficiently high stock price. A CEO is regarded overconfident if he persistently exercises options later than suggested by the benchmark; he is overconfident in his ability to keep the company's stock price rising and that he wants to profit from expected price increases by holding the options. In this regard, Holder 67 adopts 67% in-the-money during the fifth year as the threshold. If an option is more than 67% in-the-money at some point in year 5, the CEO should have exercised at least some portion of the package during or before the fifth year. This threshold corresponds to a risk-aversion measure of three in a constant relative risk-aversion utility specification. The first instance, if any, is then identified, at which the CEO failed to exercise such an option during or before the fifth year. From this point in time onward, the CEO is classified as overconfident if he subsequently exhibits the same

Continued

---

behaviour at least one more time during his tenure as CEO. To the extreme, if a CEO is optimistic enough about his firm's future performance that he holds options all the way to expiration, he is regarded overconfident as Longholder. The last measure, Net Buyer, classifies CEOs who habitually increase their holdings of their company's stock as overconfident. More precisely, a CEO is identified as overconfident if he was net buyer of his company's stock during his first five years in the sample.

The data sample used in this empirical study consists of 477 large publicly traded US firms from 1980 to 1994. The included firm must appear at least four times on one of the lists of the largest US companies compiled by the Forbes magazine in the period from 1984 to 1994. The data set provides detailed information on the stock ownership and set of option packages for the CEO of each company, year by year. The data set is supplemented by Compustat and other databases for related information on company and CEO profiles. This is a straightforward panel of time series and cross-sectional data set. Due to some restrictions and deletions with reasons given in the paper, the number of observations is 1058 for Holder 67, 3742 for Longholder and 842 for Net Buyer.

The empirical part of the study runs regression of investment on cash flow, market value of assets over book value of assets, a measure of overconfidence and a set of controlling variables. These controlling variables include corporate governance, stock ownership, and vested options. Cash flow is earnings before extraordinary items plus depreciation and is normalised by capital at the beginning of the year. The market value of assets over the book value of assets at the beginning of the year is represented by Q. Holder 67, Longholder and Net Buyer are dummies, taking the value of one if the CEO is classified as being overconfident and zero otherwise by the respective measure. Stock ownership is the fraction of company stock owned by the CEO and his immediate family at the beginning of the year. Vested options are the CEO's holdings of options that are exercisable within 6 months of the beginning of the year, as a fraction of common shares outstanding. Vested options are multiplied by 10 so that the mean is comparable to stock ownership. Size is the natural logarithm of assets at the beginning of the year. Corporate governance is the number of outside directors who currently serve as CEOs of other companies. Industries are defined as the 12 Fama-French industry groups. The dependent variable in the regression is Investment, defined as firm capital expenditures and normalised by capital at the beginning of the year. The study adopts the panel data model with fixed effects in its empirical analysis, with fixed time effects, fixed firm effects and fixed industry effects. The main results with regard to the effect of CEO overconfidence on investment that are related to panel data analysis are summarised and reported in Table 13.1. The coefficients of the interaction

Table 13.1 Regression of investment on cash flow and overconfidence

	Holder 67	Longholder	Net buyer
Cash-flow	1.7044	0.656	1.555
	(10.20)***	(7.50)***	(6.99)***
Q	−0.0088	0.0851	0.0770
	(0.44)	(7.89)***	(3.57)***
Stock-ownership (%)	−0.1834	0.196	−0.0964
	(0.33)	(2.41)**	(0.24)
Vested-options	0.1398	0.003	0.0639
	(1.17)	(0.03)	(0.42)
Size	0.0543	−0.0494	−0.0790
	(2.88)***	(5.12)***	(3.12)***
Corporate-governance	−0.0071	0.0023	0.0071
	(0.92)	(0.59)	(0.74)
$(Q)\times$ (Cash-flow)	0.0648	−0.0099	−0.0721
	(3.28)***	(1.02)	(3.17)***
(Stock-ownership) $\times$ (Cash-flow)	−0.6897	0.002	0.3991
	(1.67)*	(0.01)	(0.56)
(Vested-options) $\times$ (Cash-flow)	−0.2981	0.2847	−0.0012
	(2.62)***	(3.97)***	(0.01)
(Size) $\times$ (Cash-flow)	−0.1754	−0.053	−0.1653
	(8.77)***	(5.04)***	(6.02)***
(Corporate-governance) $\times$ (Cash-flow)	0.0441	−0.0096	0.0006
	(2.65)***	(1.07)	(0.03)
Holder 67	−0.0495		
	(1.96)*		
(Holder 67) $\times$ (Cash-flow)	0.2339		
	(4.70)***		
Longholder		−0.0504	
		(2.65)***	
(Longholder) $\times$ (Cash-flow)		0.1778	
		(5.51)***	
Net-buyer			1.0615
			(2.83)***
(Net-buyer) $\times$ (Cash-flow)			0.4226
			(4.33)***
Year-fixed-effects	Yes	Yes	Yes
Firm-fixed-effects	Yes	Yes	Yes
(Year-fixed-effects) $\times$ (Cash-flow)	Yes	Yes	Yes
(Industry-fixed-effects) $\times$ (Cash-flow)	No	No	No
(Firm-fixed-effects) $\times$ (Cash-flow)	No	No	No
Observations	1058	3742	842
Adjusted R^2	0.62	0.54	0.54

Constant included.
t-statistic in parentheses.
*significant at the 10 per cent level; **significant at the 5 per cent level; ***significant at the 1 per cent level.

Continued

of all the overconfidence measures with cash flow are significantly positive. The authors therefore conclude that overconfident CEOs have higher sensitivity to investment to cash flow than their peers. In addition, there are variations among firms and over time, as well as the interaction of fixed time effects and cash flow.

Example 13.2

Dividend policy is one of the areas in corporate finance that attracts heated debate. Using a panel of 330 large listed firms in the UK sampled in the period of 1985–1997, Khan (2006) investigates the effect of the ownership structure of firms on their dividend policies. The relationship between dividend payouts and ownership structure can be explained by agency theory that dividend payouts help mitigate the conflict of interest between a firm's management and its shareholders. Dividend provides indirect control benefits in the absence of active monitoring of firms' management by its shareholders. Dividend payouts tend to be higher if ownership is dispersed.

To empirically investigate the relationship between dividend payouts and ownership structure as suggested by the pertinent theories, a dynamic panel data model is applied in this study where the dependent variables included in the model consist of sales, net profits, financial leverage, ownership measured as at the end of the previous financial year, and dividends in the previous financial year. Ownership is represented by these variables: TOP5, the proportion of equity held by the largest five shareholders; INS, the proportion of equity owned by insurance companies; IND, the proportion of equity owned by individuals; INS5+%, the sum of equity in blocks larger than 5 per cent owned by insurance companies; and IND5+%, the sum of equity in blocks larger than 5 per cent owned by individuals. The dependent variable is dividends. All the variables are scaled by sales. This is a dynamic panel data model, since the lagged dependent variable is included as one of the regressors. The system GMM is used in its econometric analysis, because OLS estimators are biased for such dynamic panel models.

Selected estimation results are presented in Table 13.2. The OLS results are reported for the purposes of comparison with GMM estimators and revelation of the bias. The main results of the study are provided by GMM 1 and GMM 2 in the table. From GMM 1, the coefficient of TOP5 is positive and statistically significant at the 10 per cent level, and the coefficient of the square of this measure, Sq-TOP5, is negative and statistically significant at the 1 per cent level. The author therefore claims that these results indicate that the relationship between dividends and ownership concentration is concave.

Table 13.2 Estimation of effects of ownership on dividend payouts

	OLS		GMM 1		GMM 2	
Lagged-dividends	0.799	(0.017)***	0.654	(0.046)***	0.682	(0.041)***
Profits	0.167	(0.012)***	0.198	(0.029)***	0.182	(0.028)***
Lagged	−0.053	(0.01)***	−0.046	(0.02)**	−0.043	(0.022)**
1/Real-sales	0.244	(0.059)***	0.123	(0.086)	0.186	(0.073)***
Lagged	−0.203	(0.053)***	−0.075	(0.078)	−0.095	(0.066)
Leverage	0.008	(0.005)*	0.011	(0.005)**	0.010	(0.006)*
Lagged	−0.01	(0.006)*	−0.010	(0.007)	−0.008	(0.008)
TOP5	0.001	(0.001)	0.0028	(0.0016)*	0.0021	(0.0015)
Lagged	−0.002	(0.001)**	−0.0041	(0.0018)**	−0.0051	(0.0017)
Sq-TOP5	−0.0002	$(4.7e^{-5})$***	−0.0002	$(7.3e^{-5})$***	−0.00016	$(6.3e^{-5})$***
Lagged	0.0002	$(4.3e^{-5})$***	0.0001	$(7.1e^{-5})$***	0.00023	$(6.8e^{-5})$***
INS			1.2724	(0.7507)***		
Lagged			−1.0964	(0.4719)**		
INS5+%					0.034	(0.012)***
Lagged					−0.030	(0.007)***
IND5+%					−0.0022	(0.0015)
Lagged					0.0026	(0.0012)**
Diagnostic-tests						
Wald(joint)	17320	$[0.000]$***	1058	$[0.000]$***	1112.	$[0.000]$***
Sargan			257.5	$[0.996]$	212.0	$[0.784]$
m1	−1.406	$[0.160]$	−3.873	$[0.000]$***	−3.893	$[0.000]$***
m2	0.9878	$[0.323]$	1.629	$[0.103]$	1.603	$[0.109]$
Number-of-firms	330		330		330	
Number-of-observations	2,370		2,370		2,370	

Standard errors in parentheses.
* significant at the 10 per cent level; ** significant at the 5 per cent level; *** significant at the 1 per cent level.

Continued

However, the author further argues that the point where the relationship between dividends and ownership concentration turns negative occurs when total shareholding by the largest five shareholders rises above 9.6 per cent, and this point lies outside the range of the variable for the overwhelming majority of firms in the sample. The author then alleges that there is a negative but non-linear relationship between dividends and ownership concentration. Moreover, the results show that the fraction of equity owned by insurance companies, INS, has a positive effect on dividend payouts. However, this effect is almost negativated by that of its lagged variable, the fraction of equity owned by insurance companies in the previous year. If, and more or less it is the case, the fraction of equity owned by insurance companies does not change much year on year, then the results can be spurious arising from multi co-linearity between these variables. GMM 2 specification considers the effect of large block holdings by insurance companies and individual investors on dividend payouts. INS5+% and IND5+%, the sum of block shareholdings greater than 5 per cent that belong to insurance companies and individuals respectively, claims the author, are found in separate regression analyses to be statistically significant with positive and negative coefficients respectively. However, the reported GMM 2 results where INS5+% and IND5+% are considered together in the same regression, along with controls for ownership concentration, exhibit that only INS5+% remains statistically significant. Based on these results and examination, the author argues that the results indicate a negative relationship between dividends and ownership concentration overall. Though, similar to the results on INS and lagged INS with the GMM 1 specification, the effect of INS5+% on dividend payouts is almost negativated by that of its lag. So the results can be spurious caused by the same problem of multi co-linearity. Further, the study has found a positive effect on dividend payouts of insurance company share ownership and a negative effect on dividend payouts of individual share ownership. While both insurance company and individual share ownership appear to contain independent information, the most informative measure seems to be shareholding by insurance companies.

Example 13.3

CEO compensation has cast much and many controversies in the news as well as in scholarly work. The issue has attracted immense attention since it is directly associated with agent-principal theory to mitigate the conflict between, and harmonise the interest of, CEOs and shareholders. Moreover, the interest of other stakeholders, employees and customers, is also affected

by the compensation CEOs receive. In a recent attempt to establish a link between CEO compensation and firm performance, Lilling (2006) employs the GMM to examine a panel data set of North American companies. The data are drawn from over 24,000 public companies in various industries from 1993 to 2005. For usual reasons of mission data and/or unavailability of data, the final sample consists of only 1,378 firms with 6,755 total observations.

The study puts forward four hypotheses, and then empirically tests these hypotheses using the dynamic panel model. These hypotheses are: (1) there is a positive link between firm performance represented by market value of the firm and CEO compensation; (2) as the size of a firm grows in terms of sales, CEO compensation increases; (3) as a CEO has an additional year of experience, his compensation increases; and (4) a CEO hired externally will be compensated more than a CEO who is hired internally. Subsequently, in empirical dynamic panel data regression, the logarithm of market value of the firm for firm performance, the logarithm of sales for firm size, CEO tenure in years and a dummy for internally hired CEOs are included as independent variables in the model. Besides, ROA is adopted as an additional measure for firm performance, as well as a gender dummy. The dependent variable is the logarithm of total CEO compensation, including salary, bonus, restricted stock grant value, stock options, and other benefits. The dependent variable at lag one is taken in as one of the regressors, so the modelling is a typical dynamic panel data approach.

The empirical results are reported in Table 13.3 and Table 13.4, with the former using lagged firm performance and the latter using contemporary firm performance in estimation. The first differenced GMM and the system GMM have produced similar results while fixed effects estimators are rather different. Analysis and discussion are based on GMM results, since fixed effects may cause bias in parameter estimates. The coefficients of lagged CEO compensation are positive and highly significant at the 1 per cent level, indicating the series of CEO compensation are fairly persistent. This implies that the lagged level variables provide weak instruments for the differences in the first differenced GMM model; hence the author alleges the system GMM approach is more suitable than the first differenced GMM in this case. According to the estimated coefficients, CEO compensation is strongly linked to the market value of the firm, and it is more responsive to the market value of the firm in the current year than that in the previous year. The coefficient estimated by the system GMM is 0.358 in Table 13.3 and greater at 0.405 in Table 13.4; and the coefficient estimated by the first differenced GMM is 0.214 in Table 13.3 and greater at 0.444 in Table 13.4, all of them significant at the 1 per cent level. These results support the statement of the first hypothesis in the paper that there

Continued

Table 13.3 CEO compensation – estimation with lagged variables

	Fixed effects		First differenced GMM		System GMM	
Ln-Total-compensation$_{t-1}$	0.024	(0.012)*	0.103	(0.027)***	0.151	(0.022)***
Ln-Market-value$_{t-1}$	0.152	(0.020)***	0.214	(0.066)***	0.358	(0.028)***
Ln-ROA$_{t-1}$	0.002	(0.001)	0.005	(0.002)***	0.001	(0.001)
CEO-tenure	−0.022	(0.005)***	−0.065	(0.026)**	−0.037	(0.021)*
CEO-tenure^2	0.000	(.000)	0.001	(.001)	0.001	(0001)
Ln-Sales	0.255	(0.029)***	0.210	(0.052)***	0.092	(0.024)***
Gender (1 if female)	−0.039	(0.142)	0.015	(0.242)	0.035	(0.196)
CEO-internal (1 if internal)	−0.163	(0.045)***	−0.256	(0.089)***	−0.129	(0.039)***
Constant	5.133	(0.209)***	3.752	(0.144)***		
Sargan-test (*P*-value)			0.288		0.120	
AR1 (*P*-value)			0.000		0.000	
AR2 (*P*-value)			0.254		0.157	
R-squared	0.196					
Observations	6755		5289		6755	
Number-of-firms	1378		1197		1378	

Standard errors in parentheses.
* significant at the 10 per cent level; ** significant at the 5 per cent level; *** significant at the 1 per cent level.

Table 13.4 CEO compensation – estimation with contemporary variables

	Fixed effects		First differenced GMM		System GMM	
Ln-Total-compensation$_{t-1}$	0.018	(0.013)	0.092	(0.027)***	0.142	(0.021)***
Ln-Market-value	0.345	(0.018)***	0.444	(0.073)***	0.405	(0.030)***
Ln-ROA	0.003	(0.001)**	−0.004	(0.003)	−0.007	(0.002)***
CEO-tenure	−0.021	(0.005)***	−0.020	(0.026)	−0.021	(0.020)
CEO-tenure^2	0.000	(0.000)	−0.001	(0.001)	0.000	(0.001)
Ln-Sales	0.123	(0.028)***	0.150	(0.050)***	0.063	(0.023)***
Gender	−0.069	(0.138)	−0.204	(0.276)	0.086	(0.186)
CEO-internal	−0.149	(0.043)***	−0.239	(0.089)***	−0.129	(0.038)***
Constant	4.648	(0.206)***	3.521	(0.147)***		
Sargan-test (*P*-value)			0.326		0.187	
AR1 (*P*-value)			0.000		0.000	
AR2 (*P*-value)			0.155		0.064	
R-squared	0.25					
Observations	6755		5289		6755	
Number-of-firms	1378		1197		1378	

Standard errors in parentheses.
* significant at the 10 per cent level; ** significant at the 5 per cent level; ***significant at the 1 per cent level.

Continued

is a positive link between firm performance represented by the market value of the firm and CEO compensation. The association between CEO compensation and the other firm performance measure, ROA, is controversial. The coefficient is negative and highly significant at the 1 per cent level by the author's preferred model of the system GMM in Table 13.4, where contemporary ROA is used, while it is significantly positive at the 1 per cent level by the first differenced GMM in Table 13.3, where lagged ROA is used. No explanations are provided and these results on ROA seem to be ignored. Tables 13.3 and 13.4 also report a strong link between CEO compensation and sales, a proxy for firm size. The coefficient is all significantly positive at the 1 per cent level in both tables, indicating the validity of the statement of the second hypothesis that CEO compensation increases with growth in sales. The third hypothesis, CEO compensation increases with CEO tenure, is modestly supported by the results. The coefficient is positive and significant at the 10 per cent level by the system GMM and at the 5 per cent level by the first differenced GMM in Table 13.3 and negative and insignificant in Table 13.4. Moreover, none of squared CEO tenure is significant in both tables. The internally promoted CEOs are less compensated than externally recruited CEOs, evidenced by a negative coefficient for the internal CEO dummy that is significant at the 1 per cent level in both tables and by both models. This supports the statement of the fourth hypothesis in the study that an externally hired CEO will be compensated more than a CEO who is hired internally.

13.6. Empirical literature

The use of panel data and application of panel data modelling have increased significantly since the first edition of this book. This is particularly evident in finance and related areas. The volume of studies and papers employing panel data has been multifold, in recognition of the advantages offered by panel data approaches as well as panel data sets themselves, and in response to the growing availability of data sets in the form of panel. Prior to the 1990s, panel data were traditionally and predominantly used in social-economic research and social science research, despite various cases in economics and finance such as studies on PPP, estimation of stocks' betas and inquiries into firm performance. Although the use of panel data in finance and financial economics may have predated that in social-economic research, it is the recent five years that have witnessed proliferating applications of panel data sets and panel data approaches in these areas.

Firm performance and firm value are of importance to shareholders, as well as all other stakeholders, including employees, bondholders, customers and the management team. Employing a panel data set consisting of 12,508 firms over

a nine-year period from 1993 to 2001, Goddard *et al*. (2005) examine the determinants of profitability of firms in the manufacturing and services sector in four European countries of Belgium, France, Italy and the UK. They find evidence of a negative effect of size on profitability and a positive effect of market share on profitability, and the effect is greater in manufacturing than in services. The relationship between a firm's financial leverage and its profitability is found to be negative, with liquidity contributing to firms' profitability. In addition to Example 13.3 on CEO compensation and firm performance, there are a few similar studies in recent years, such as Kato and Kubo (2006) and Kato *et al*. (2007), among others. Kato and Kubo (2006) pointedly spell out that prior studies on Japanese executive compensation have been constrained by the lack of longitudinal data on individual CEO pay. Using a panel data set on individual CEOs' salary and bonus of Japanese firms spanning 10 years from 1986 to 1995, they study the pay-performance relationship for Japanese CEO compensation. It is documented that Japanese CEO cash compensation is sensitive to firm performance, especially firm performance in terms of accounting measures, and the findings are claimed to be consistent. This implies that, to a certain extent, stock market performance tends to play a less important role in the determination of Japanese CEO compensation. They also find that the bonus system makes CEO compensation more responsive to firm performance in Japan, in contrast to the argument in the literature on compensation that bonuses are disguised wages/salaries. Using a panel data set with 246 publicly traded firms in Korea from 1998 to 2001, Kato *et al*. (2007) investigate the link between executive compensation and firm performance in Korea. They find that cash compensation of Korean executives is significantly related to stock market performance. This is in contrast to the findings on Japanese CEO compensation in Kato and Kubo (2006), where CEO compensation is linked to accounting measures of firm performance consistently but it is linked to stock market performance to a lesser extent. The findings in these two studies suggest that there is a difference between Japan and Korea in CEO compensation. Nevertheless, Kato *et al*. (2007) find that the magnitude of the pay-performance sensitivity in Korea is comparable to that of the US and Japan. Further analysis is claimed to reveal that such significant executive pay-performance link is embedded with non-Chaebol firms and no such link exists for Chaebol firms. The authors therefore argue for corporate governance reforms in Korea, aiming primarily at Chaebol firms. Bhabra (2007) investigates the relationship between insider stock ownership and firm value, using a panel of 54 publicly traded firms in New Zealand for the period 1994–1998. In order to limit the effect of outliers, firms with a Tobin's Q greater than 4.00 are excluded.

Also excluded are firms with insider ownership being lower than 0.1 per cent. It is alleged that insider ownership and firm value are positively related for ownership levels below 14 per cent and above 40 per cent and inversely related at intermediate levels of ownership. These results are fairly robust to different measures of firm performance, including Tobin's Q, market to book ratio and ROE and to several different estimation techniques. Employing a panel of non-financial companies that trade on the Spanish Continuous Market for the

period 1998–2000, Minguez-Vera *et al*. (2007) evaluate the effect of ownership structure on firm value. It is claimed to have found a positive effect of the ownership by major shareholders on firm value and a positive effect of the degree of control on Tobin's Q. However, relationship between the ownership of large block shareholders and firm value is insignificant; a positive effect is identified when the major shareholders are individuals.

Foreign direct investment (FDI) has been an intensive activity over the last three decades, with significant effects on firms at the micro level as well as on the national economy at the macro level, in an accelerating process of globalisation and amongst national economies and ever-interwoven multi-national companies. Cross-country cases themselves constitute cross-sectional data sets that conveniently become panel data over time, which provides rich sources for empirical studies as well as develops motivations for model enhancement. Employing a panel data set for 22 countries over the period 1984–2000, Asiedu (2006) examines the impact of natural resources, market size, government policies, political instability and the quality of the host country's institutions on FDI. It is found that natural resources and large markets promote FDI, along with lower inflation, good infrastructure, an educated population, openness to FDI, less corruption, political stability and a reliable legal system. The author alleges that a decline in corruption from the level of Nigeria to that of South Africa has the same positive effect on FDI as increasing the share of fuels and minerals in total exports by about 35 per cent, suggesting that countries that are small or lack natural resources can also attract FDI by improving their institutions. Naude and Krugell (2007) use a panel of 43 African countries for the period 1970–1990 in an empirical study. They identify government consumption, inflation, investment, political stability, accountability, regulatory burden and rule of law, and initial literacy as the determinants of FDI in Africa. It is claimed that geography does not seem to have a direct influence on FDI flows to Africa. Hansen and Rand (2006) use a panel data set for 31 developing countries over 31 years to conduct Granger causality tests for FDI and growth relationships. They find that FDI has a lasting impact on GDP, while GDP has no long-run impact on the FDI-to-GDP ratio. Based on this finding, they conclude that FDI causes growth. Similarly, Eller *et al*. (2006) examine the impact of financial sector FDI on economic growth via the efficiency channel, employing a panel data set for 11 Central and Eastern European countries for the period 1996–2003. They claim to have found a hump-shaped impact of financial sector FDI on economic growth. Using a panel data set for eight Asian countries and regions of China, Korea, Taiwan, Hong Kong, Singapore, Malaysia, Philippines and Thailand for the period 1986–2004, Hsiao and Hsiao (2006) test Granger causalities between GDP, exports and FDI. The Granger causality test results indicate that FDI has unidirectional effects on GDP directly and also indirectly through exports. In addition, there exist bidirectional causalities between exports and GDP. Investigating whether FDI is a channel of technology transfer and knowledge spillovers, Bwalya (2006) uses firm-level data from Zambia to analyse the nature and significance of productivity externalities of FDI to local firms. The data set used in the study includes 125 firms in sectors of food, textile, wood and metal and is obtained from the World Bank

through the Regional Program on Enterprise Development survey conducted in 1993, 1994 and 1995. A firm is classified as 'foreign' if it has at least 5 per cent foreign shareholding in it. It is claimed that significant inter-industry knowledge spillovers have been found through linkages, whereas there is little evidence in support of intra-industry productivity spillovers from FDI. The net impact of FDI depends on the interaction between intra-industry and inter-industry productivity effects. Kostevc *et al*. (2007) pay attention to the relationship between FDI and the quality of the institutional environment in transition economies. Using a panel data set of 24 transition economies for the period 1995–2002, they assert that there is significant impact of various institutional characteristics on the inflow of foreign capital. It is found that the quality of the institutional environment has significantly influenced the level of FDI in transition economies in the observed period.

The study by Kimino *et al*. (2007) is different in that it examines FDI flows into Japan, in contrast to most studies in the area where developing countries are on the receiving side of FDI. The data used in the study consist of a panel of 17 source countries for the period 1989–2002. Six hypotheses have been tested: (1) there is a positive relationship between the market size of source countries and FDI inflows to Japan; (2) there is a negative relationship between source country exports and FDI inflows to Japan; (3) there is a positive relationship between appreciation of the source country currency and FDI inflows to Japan; (4) there is a negative relationship between the cost of borrowing differentials and FDI inflows to Japan; (5) there is a positive relationship between labour cost differentials and FDI inflows to Japan; and (6) there is a positive relationship between the investment climate of source countries and FDI inflows to Japan. Accepting the null of the second hypothesis, their results suggest that FDI into Japan is inversely related to trade flows, such that trade and FDI are substitutes. It is found that FDI increases with home country political and economic stability, confirming the statement of the sixth hypothesis. The authors argue that results regarding exchange rates, relative borrowing costs and labour costs in explaining FDI flows are sensitive to econometric specifications and estimation approaches, based on the testing results on the third, fourth and fifth hypotheses. The paper plays down the barriers to inward investment penetration in Japan and the negative effect of cultural and geographic distance. Meanwhile, it emphasises that the attitude to risk in the source country is a major factor strongly related to the size of FDI flows to Japan.

Branstetter (2006) scrutinises whether FDI is an important channel for the mediation of knowledge spillovers by analysing international knowledge spillovers originated from Japan's FDI in the US at the firm level. The study uses patent citation data to infer knowledge spillovers and empirical analysis is based on a panel data set for 189 Japanese firms for the years 1980–1997. Data on FDI are compiled from Firms Overseas Investment 1997 and 1999 editions, while patent data are obtained from the US Patent and Trademark Office and the NBER Patent Citation database. Testing the hypothesis that FDI is a channel of knowledge spillovers for Japanese multinationals undertaking direct investments in the US, the paper has found evidence that FDI increases the flow of knowledge spillovers

both from and to the investing Japanese firms. Using longitudinal panel data on Japanese firms for the years 1994–2000, Kimura and Kiyota (2006) examine the relationship between exports, FDI and firm productivity. The longitudinal panel data set is compiled through making use of the micro database of the Results of the Basic Survey of Japanese Business Structure and Activities by Research and Statistics Department, Ministry of Economy, Trade and Industry. The number of firms included amounts to over 22,000 each year. The main findings are straightforward: the most productive firms engage in exports and FDI, medium productive firms engage in either exports or FDI, and the least productive firms focus only on the domestic market. It has been found that exports and FDI appear to improve firm productivity once the productivity convergence effect is controlled for. Firms that retain a presence in foreign markets either by exports or FDI show the highest productivity growth. This overall contributes to improvements in national productivity. In an attempt to contribute to the debate as to whether outward FDI complements or substitutes for a home country's exports, Seo and Suh (2006) investigate the experience of Korean outward FDI in the four ASEAN countries of Indonesia, Malaysia, the Philippines, and Thailand during the 1987–2002 period. It is found that contemporaneous FDI flows marginally contribute to Korea's exports to the four ASEAN countries, but FDI stocks in these countries do not have discernible trade substituting effects on either exports or imports by Korea. Todo (2006) investigates knowledge spillovers from FDI in R&D, using firm-level panel data from the Japanese manufacturing industries for the period 1995–2002. Data are collected from Basic Survey of Enterprise Activities by Ministry of Economy, Trade and Industry that covers all Japanese firms in manufacturing industries that employ 50 employees or more. To distinguish between spillovers from production and R&D activities of foreign firms, the empirical study estimates the effect of each of physical capital and R&D stocks of foreign firms in the industry on the total factor productivity (TFP) level of domestic firms in the same industry. The analysis has found positive effects of R&D stocks of foreign firms on the productivity of domestic firms when the effects of capital stocks of foreign firms are absent. The author therefore alleges that knowledge of foreign firms spills over through their R&D channels, not through their production activities. Moreover, the extent of spillovers from R&D stocks of foreign firms is substantially greater than that of spillovers from R&D stocks of domestic firms. Barrios *et al.* (2005) examine the impact of FDI on the development of local firms by focusing on two likely effects of FDI: a competition effect which deters entry of domestic firms and positive market externalities which foster the development of the local industry. They argue that the competition effect first dominates but it is gradually outweighed by positive externalities. Their plant level data span the period 1972–2000 and are from the source of the annual Employment Survey collected by Forfás, the policy and advisory board for industrial development in Ireland. Since the response rate to this survey is estimated by Forfás to be generally well over 99 per cent, they claim that their data set can be seen as including virtually the whole population of manufacturing firms in Ireland. It is claimed that evidence for Ireland tends to support this result. For the manufacturing

sector, it is found that the competition effect may have deterred local firms' entry initially. However, this initial effect of competition is later outweighed by the effect of positive externalities. Overall the impact of FDI is largely positive for the domestic industry.

PPP has been one of the ever-lasting topics subjected to intensive scrutiny in the history of economic thought. There is a slight difference in the construction between the panel of PPP and, say, the panel for firm performance and its determinants. Here we usually have the time series of individual pairs of countries and the exchange rates between their currencies first; then we pool them across sections; and the estimation power may be enhanced not only for taking in more information and observations but also for taking into account potential cross-section relations and interactions. Whereas, for the study of firm performance, cross-sections come first; and then the estimation is enhanced by pooling them over time, with time patterns being identified in the meantime. PPP is important as a theory in economics; it is as well relevant to corporate decisions and strategies of international dimensions where a sound assessment of the economic environment in which the firm is to operate is crucial to success. Tests for PPP empirically amount to testing cointegration between the price levels in a pair of countries and the exchange rate between their currencies; or testing stationarity of the real exchange rate that is the nominal exchange rate adjusted by two countries' price levels. Using an annual sample of 21 OECD country CPI-based real exchange rates from 1973 to 1998, and controlling for multiple sources of bias, Choi *et al.* (2006) estimate the half-life to be 3 years with a 95 per cent confidence interval ranging from 2.3 to 4.2 years. That is, it takes three years for a deviation from PPP to fall to its half size. The shorter the half-life, the speedier is the convergence to PPP following some divergence. A sample with 26 annual observations certainly falls short of the basic requirements for statistical estimation. Increasing the number of data points by combining the cross-section with the time series should give more precise estimates, as the authors pointedly stress in the paper, which yields over 500 observations for their empirical study. Murray and Papell (2005) test for PPP with a panel of 20 countries' quarterly real exchange rates *vis-à-vis* the US dollar, from the first quarter in 1973 to the second quarter in 1998, when the nominal exchange rates between the countries in the euro zone became irreversibly fixed. The 20 countries included in the study are: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland and the UK. They assert that half-life remains 3–5 years for the post-1973 period following the float of exchange rates. That a flexible exchange rate regime does not expedite PPP convergence seems puzzling, mystifying the PPP puzzle that has already existed so long. Harris *et al.* (2005) investigate PPP for a group of 17 countries using a panel based test of stationarity that allows for arbitrary cross-sectional dependence. They use monthly data of real exchange rates of 17 countries of Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Japan, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the UK, *vis-à-vis* the US dollar between 1973: 01 and 1998: 12. They treat the short-run

time series dynamics non-parametrically to avoid the need to fit separate, and potentially mis-specified, models for the individual series. It is documented that significant evidence is found against the PPP null hypothesis even when allowance is made for structural breaks. Testing for unit roots in real exchange rates of 84 countries against the US dollar, presumably for varied periods, Alba and Papell (2007) examine long-run PPP. They claim to have found stronger evidence of PPP in countries more open to trade, closer to the US, with lower inflation and moderate nominal exchange rate volatility, and with similar economic growth rates as the US. It is also observed that PPP holds for panels of European and Latin American countries, but not for African and Asian countries. Based on the above results, the authors conclude that their findings demonstrate that country characteristics can help explain both adherence to and deviations from long-run PPP. Questioning the unit root test results of PPP in previous studies adopting panel methods, Banerjee *et al*. (2005) offer the reasons why PPP usually holds when tested for in panel data but usually does not hold in univariate analysis. They challenge the usual explanation for this mismatch that panel tests for unit roots are more powerful than their univariate counterparts. They demonstrate that the assumption of non-presence of cross-section cointegration relationships in existing panel methods is dubious. Cross-section cointegration relationships would tie the units of the panel together, which tends to make the test results appear stationary. Using simulations, they show that if this important underlying assumption of panel unit root tests is violated, the empirical size of the tests is substantially higher than the nominal level, and the null hypothesis of a unit root is rejected too often even when it is true. Subsequently, they warn against the 'automatic' use of panel methods for testing for unit roots in macroeconomic time series, in addition to testing for PPP.

There appear to be numerous subjects and abundant studies and papers where panel data and panel data methods are employed in recent three–five years. Nier and Baumann (2006) study the factors that influence banks to limit their risk of default. They have constructed and used a panel data set consisting of observations on 729 individual banks from 32 different countries over the period 1993–2000 in their empirical estimation. Their results suggest that government safety nets result in lower capital buffers and that stronger market discipline stemming from uninsured liabilities and disclosure results in larger capital buffers, holding other factors constant. It is found that the effect of disclosure and uninsured funding is reduced when banks enjoy a high degree of government support. Following certain deliberation, they claim that, while competition leads to greater risk taking incentives, market discipline is more effective in curbing these incentives in countries where competition among banks is strong. On firms' risk management practice, Fehle and Tsyplakov (2005) propose an infinite-horizon, continuous-time model of a firm that can dynamically adjust the use of risk management instruments. The model aims at reducing uncertainties in product prices and thereby mitigating financial distress losses and reducing taxes. In the model, the firm can adjust its use, the hedge ratio and maturity of risk management instruments over time, and transaction costs are associated with initiation and adjustment of risk management contracts. They claim that the model produces a

number of new time-series and cross-sectional implications on how firms use short-term instruments to hedge long-term cash flow uncertainties. They then use quarterly panel data on gold mining firms between 1993 and 1999 to fit the proposed model. A non-monotonic relation between measures of financial distress and risk management activity is found with the panel data set, claimed to be consistent with the model. The importance of R&D in business has been gradually acknowledged over time. To answer what drives business R&D, Falk (2006) uses a panel of OECD countries for the period 1975–2002 with data measured as five-year averages. Korea, Mexico, Czech Republic and Hungary are excluded since they joined OECD fairly recently. It is found that tax incentives for R&D have a significantly positive impact on business R&D expenditure, regardless of the specification and estimation techniques. R&D expenditure by universities is found to be significantly positively related to business sector R&D expenditure, and the author claims that this result indicates that public sector R&D and private R&D are complementary to each other. It is also found that direct R&D subsidies and high-tech export shares are significantly positively related to business-sector R&D intensity to a certain extent. Venture capital (VC) has attracted much media attention and attention of academics recently. Schertler (2007) investigate whether countries with high knowledge capital show higher volumes of VC investments than countries with low knowledge capital. To this end, a panel data set of 15 European countries over the years 1991–2001 is used. Included variables in this panel data set enable the model to test the impact of previous years' knowledge capital on the volume of today's VC investments. There is evidence that the measure of total knowledge capital, which is the sum of government-financed and business-financed knowledge capital, has strong explanatory power for VC investments. The result is robust with regard to various measures of knowledge capital, such as the number of patent applications, the number of R&D researchers or gross expenditures on R&D. In addition, weak evidence is found that the measure of government-financed knowledge capital has a positive effect on VC investments with a delay of several years.

Questions and problems

1 What is defined as panel data? How do cross-sectional data and longitudinal data/time series data form a panel data set?

2 Present and describe several commonly used panel data resulting from surveys in social-economic research, such as The National Longitudinal Surveys of Labor Market Experience of the US.

3 How do data samples used for testing PPP constitute panel data? What are the advantages of adopting panel data approaches to examining PPP?

4 Why is it claimed that the use of panel data in finance and financial economics predated that in social-economic research, and the use is very extensive too? Provide a few of examples.

5 Describe and contrast stacked and unstacked panel data structures in the organisation and presentation of panel data.

6 What are fixed effects in panel data modelling? What are random effects in panel data modelling? Contrast them with each other with respect to the assumptions on the residual's correlation structure.

7 What are featured by random parameter or random coefficient models in panel data modelling? Contrast random parameter models with random effects models.

8 What is meant by dynamic panel data analysis? What kinds of issues may arise in performing dynamic panel data analysis?

9 Describe and discuss commonly applied approaches to estimating dynamic panel data models.

10 Collect data on variables related to the determination of corporate capital structure from FAME or Thomson ONE Banker and company annual reports. Construct a panel data set for N firms over a T year period with K independent variables. Build a panel data model where the dependent variable is the debt-to-equity ratio or financial leverage (using EViews, LIMDEP, RATS or other packages). Firm size, profitability, profit variability, business risk, growth opportunities, non-debt tax shields, cash holdings, dividend payouts, collateral assets, asset tangibility, uniqueness, shareholder structure and concentration, and an industry dummy may be considered for independent variables.

11 Repeat the above with the lagged debt-to-equity ratio or financial leverage being included as an additional regressor. The model becomes dynamic so pertinent approaches and procedures should be followed.

12 Collect data on inward FDI at the country level for N host countries over a T year period and with K independent variables. Build and then estimate a panel data model (using EViews, LIMDEP, RATS or other packages). The dependent variable is inward FDI flows to the host country (with necessary transformation/scaling). The independent variables may include host country infrastructure, relative labour costs, changes in, and stability of, exchange rates, political stability and legal systems

13 Repeat the above with the lagged inward FDI flows being included as an additional regressor. Apply pertinent approaches and procedures to estimate this dynamic panel data model.

14 Collect foreign exchange spot rates and the corresponding forward rates for several currencies *vis-à-vis* the euro over a certain period. Build a panel data model to test the hypothesis that the forward premium is an unbiased predictor of future spot rate changes.

References

Alba, J.D. and Papell, D.H. (2007), Purchasing power parity and country characteristics: evidence from panel data tests, *Journal of Development Economics*, **83**, 240–251.

Arellano, M. and Bond, S.R. (1991), Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies*, **58**, 277–297.

Arellano, M. and Bover, O. (1995), Another look at the instrumental-variable estimation of error components models, *Journal of Econometrics*, **68**, 29–52.

Asiedu, E. (2006), Foreign direct investment in Africa: the role of natural resources, market size, government policy, institutions and political instability, *World Economy*, **29**, 63–77.

Banerjee, A., Marcellino, M. and Osbat, C. (2005), Testing for PPP: should we use panel methods? *Empirical Economics*, **30**, 77–91.

Barrios, S., Gorg, H. and Strobl, E. (2005), Foreign direct investment, competition and industrial development in the host country, *European Economic Review*, **49**, 1761–1784.

Bhabra, G.S. (2007), Insider ownership and firm value in New Zealand, *Journal of Multinational Financial Management*, **17**, 142–154.

Blundell, R. and Bond, S. (1998), Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics*, **87**, 115–143

Branstetter, L. (2006), Is foreign direct investment a channel of knowledge spillovers? evidence from Japan's FDI in the United States, *Journal of International Economics*, **68**, 325–344.

Bwalya, S.M. (2006), Foreign direct investment and technology spillovers: evidence from panel data analysis of manufacturing firms in Zambia, *Journal of Development Economics*, **81**, 514–526.

Choi, C.Y, Mark, N.C. and Sul, D. (2006), Unbiased estimation of the half-life to PPP convergence in panel data, *Journal of Money, Credit, and Banking*, **38**, 921–938.

Eller, M., Haiss, P. and Steiner, K. (2006), Foreign direct investment in the financial sector and economic growth in Central and Eastern Europe: the crucial role of the efficiency channel, *Emerging Markets Review*, **7**, 300–319.

Falk, M. (2006), What drives business research and development (R&D) intensity across organisation for Economic Co-operation and Development (OECD) countries? *Applied Economics*, **38**, 533–547.

Fehle, F. and Tsyplakov, S. (2005), Dynamic risk management: theory and evidence, *Journal of Financial Economics*, **78**, 3–47.

Goddard, J., Tavakoli, M. and Wilson, J.O.S. (2005), Determinants of profitability in European manufacturing and services: evidence from a dynamic panel model, *Applied Financial Economics*, **15**, 1269–1282.

Hansen, H. and Rand, J. (2006), On the causal links between FDI and growth in developing countries, *World Economy*, **29**, 21–41.

Harris, D., Leybourne, S. and McCabe, B. (2005), Panel stationarity tests for purchasing power parity with cross-sectional dependence, *Journal of Business and Economic Statistics*, **23**, 395–409.

Hsiao, F.S.T. and Hsiao, M.C.W. (2006), FDI, exports, and GDP in East and Southeast Asia – panel data versus time-series causality analyses, *Journal of Asian Economics*, **17**, 1082–1106.

Kato, T. and Kubo, K. (2006), CEO compensation and firm performance in Japan: evidence from new panel data on individual CEO pay, *Journal of the Japanese and International Economies*, **20**, 1–19.

Kato, T., Kim, W. and Lee, J.H. (2007), Executive compensation, firm performance, and Chaebols in Korea: evidence from new panel data, *Pacific-Basin Finance Journal*, **15**, 36–55.

Khan, T. (2006), Company dividends and ownership structure: evidence from UK panel data, *Economic Journal*, **116**, C172–C189.

Kimino, S., Saal, D.S. and Driffield, N. (2007), Macro determinants of FDI inflows to Japan: an analysis of source country characteristics, *World Economy*, **30**, 446–469.

Kimura, F. and Kiyota, K. (2006), Exports, FDI, and productivity: dynamic evidence from Japanese firms, *Review of World Economics/Weltwirtschaftliches*, **142**, 695–719.

Kostevc, C., Redek, T. and Susjan, A. (2007), Foreign direct investment and institutional environment in transition economies, *Transition Studies Review*, **14**, 40–54.

Lilling, M.S. (2006), The link between CEO compensation and firm performance: does simultaneity matter? *Atlantic Economic Journal*, **34**, 101–114.

Malmendier, U. and Tate, G. (2005), CEO overconfidence and corporate investment, *Journal of Finance*, **60**, 2661–2700.

Minguez-Vera, A. and Martin-Ugedo, J.F. (2007), Does ownership structure affect value? a panel data analysis for the Spanish market, *International Review of Financial Analysis*, **16**, 81–98.

Murray, C.J. and Papell, D.H. (2005), Do panels help solve the purchasing power parity puzzle? *Journal of Business and Economic Statistics*, **23**, 410–415.

Naude, W.A. and Krugell, W.F. (2007), Investigating geography and institutions as determinants of foreign direct investment in Africa using panel data, *Applied Economics*, **39**, 1223–1233.

Nickell, S.J. (1981), Biases in dynamic models with fixed effects, *Econometrica*, **49**, 1417–1426.

Nier, E. and Baumann, U. (2006), Market discipline, disclosure and moral hazard in banking, *Journal of Financial Intermediation*, **15**, 332–361.

Schertler, A. (2007), Knowledge capital and venture capital investments: new evidence from European panel data, *German Economic Review*, **8**, 64–88.

Seo, J.S. and Suh, C.S. (2006), An analysis of home country trade effects of outward foreign direct investment: the Korean experience with ASEAN, 1987–2002, *ASEAN Economic Bulletin*, **23**, 160–170.

Todo, Y. (2006), Knowledge spillovers from foreign direct investment in R&D: evidence from Japanese firm-level data, *Journal of Asian Economics*, **17**, 996–1013.

14 Research tools and sources of information

This chapter is intended to help the reader carry out an empirical modern financial economics or econometrics project. The chapter recommends relevant on-line information and literature on research in financial markets and financial economics. Some commonly used econometrics software packages for time series analysis, as well as cross-sectional and panel data analysis, are introduced. We feel that perfection of an empirical study can only be achieved against a wider background of the business environment, market operations and institutional roles, and by frequently upgrading the knowledge base. To this end, the coverage of this chapter is extended to include major monetary and financial institutions, international organisations, stock exchanges and option and futures exchanges, and professional associations and learned societies. Most materials are in the form of websites, which can be accessed almost instantly from anywhere in the world. In doing so, this chapter does not only endow the reader with various tools and information for empirical research, but also prompts and/or reminds the researcher of the factors and players to be considered in the research.

14.1. Financial economics and econometrics literature on the Internet

Mostly and increasingly, finance journals are covered by lists of economics journals on the web. The following two sites are comprehensive and frequently used by academia and professionals alike:

http://www.oswego.edu/~economic/journals.htm

at the State University of New York (SUNY), Oswego; and

http://www.helsinki.fi/WebEc

at University of Helsinki, Finland. These sites provide editorial information, tables of contents and abstracts for most of the listed journals. To make access to full

papers, one has to contact the publisher. More and more, there are Internet journal archive service agencies, one of the most influential is:

http://www.jstor.org

For major finance journals, it is worthwhile visiting:

http://www.cob.ohio-state.edu/fin/journal/jofsites.htm#otjnl

at the Ohio State University. Indeed, the Ohio State University maintains wide-ranging financial sites:

http://www.cob.ohio-state.edu/fin/journal/jofsites.htm

including Finance Journals, Institutional Working Paper Sites, Personal Working Paper Sites, The Finance Profession, Research Centers, Link Collections, Asset Pricing and Investments, Derivatives, Corporate Finance and Governance, Financial Institutions, Research Software and Data, Educational Resources, Of Interest to Students, and Miscellanies.

Not only economics journals, but also finance journals, classify paper topics by the JEL (*Journal of Economic Literature*) Classification System. The JEL classification numbers are now also available on the web:

http://www.aeaweb.org/journal/elclasjn.html

Social Science Research Network (SSRN) is very active in disseminating research output. Its website is:

http://www.ssrn.com

SSRN consists of five sub-networks: Accounting Research Network (ARN), Economics Research Network (ERN), Latin American Network (LAN), Financial Economics Network (FEN), and Legal Scholarship Network (LSN). The most relevant networks for the topics in this book are FEN and ERN. SSRN publishes working papers and abstracts of journal papers, downloadable free of charge. It encourages scholars to electronically submit their working papers and the abstracts and rank the papers by download, so it constitutes an efficient channel for gathering information on the most recent developments in the areas.

Other useful sites include:

Resources for Economists on the Internet

http://www.rfe.org

This site is sponsored by the American Economic Association and maintained by the Economics Department at the State University of New York, Oswego. It lists more than 2,000 resources in 97 areas available on the Internet, with reasonable descriptions. Covered are organisations and associations, data and software, mailing lists and forums, meetings, conferences and announcements,

and more. Its Economics Search Engine is helpful for economists to search the Internet for economic information. There is even an area for blogs and commentaries.

CRSP

http://www.crsp.com

CRSP stands for Center for Research in Securities Prices that was founded in 1960. It is a research centre of Chicago Graduate School of Business. CRSP files cover common stocks listed on the NYSE, AMEX and NASDAQ stock markets, US Government Treasury issues, and US mutual funds. CRSP has a wide variety of financial and economic indices (market, total return, cap-based and custom) and other statistics used to gauge the performance of the broader market and economy in general. CRSP also provides proxy graphs for 10K SEC filing, monthly cap-based reports and custom data sets and extractions.

Mimas

http://www.mimas.ac.uk

Mimas is a JISC and ESRC-supported national data centre run by Manchester Computing, at the University of Manchester. It provides the UK higher education, further education and research community with networked access to key data and information resources – socio-economic, spatial and scientific data, and bibliographic – to support teaching, learning and research across a wide range of disciplines. Mimas services are available free of charge to users at eligible institutions.

Econometric Links

http://econometriclinks.com

Maintained by *The Econometrics Journal*, the site covers software, codes, data and other sections and the links are wide-ranged. For example, it links to numerous codes for a variety of software packages, including RATS, GAUSS, Matlab, Mathematica, SHAZAM, Fortran and C++. The site also brings together various data sources such as Bureau of Economic Analysis data, Morgan Stanley Capital International, Datasets from the National Bureau of Economic Research, and Dow Jones indexes. A rich list of journals in statistics and econometrics and their web links are provided, as is information on econometrics conferences and workshops.

14.2. Econometric software packages for financial and economic data analysis

None of modern financial econometrics projects can be executed without making use of an econometrics package. Following is a list of popular contemporary

packages being widely, but not exclusively, used by financial economists and econometricians.

EViews

http://www.eviews.com

EViews, or Econometric Views, is a menu-driven and user-friendly package. The most recent version is EViews 6. It can easily handle most modern econometric models such as binary dependent variable models, univariate GARCH, cross section and panel data, and so on. Its help system in the electronic form is excellent, e.g., the Estimation Methods part provides detailed information on model specification and estimation, as well as the background and origin of the model. However, the menu-driven feature also means that the package is not flexible to adapt to the need of specific requirements. Although there are many variations of GARCH available, they are all univariate. The state space model and the Kalman filter can only do basic things, which are far from enough to cope with the requirement encountered in modern empirical studies featured by sophisticated model specifications and extensions. For more detail and purchase information visit the website given.

RATS

http://www.estima.com

RATS, Regressional Analysis of Time Series, is one of the most authoritative packages in the area. With RATS version 7 there are many new features and improvements over the previous versions, such as more instrument variable/GMM support, and improved graphics. Like EViews, RATS also has User's Guide and Reference Manual in the electronic form, though it is an industry norm now. One of the advantages for using RATS is that, while a specialist package for time series analysis equipped with many readily executable procedures, the user can write or easily adapt a procedure for her/his own specific needs, or s/he can even write a procedure from scratch. Therefore, even if GARCH procedures were not provided, the user can write one with, e.g., RATS functions and the maximum likelihood procedure. As such, virtually all kinds of contemporary time series models can be estimated with RATS, though sometimes it involves great complexity and requires much experience and skill. Mainly a time series package, one can also programme models of cross section and panel data with RATS. In addition to conventional analysis in the time domain, RATS can estimate time series in the frequency domain, also known as spectral analysis of time series. Spectral analysis with RATS includes the Fourier transform, spectra and cross spectra, and coherence and phase. All of these are at the application level capable of handling empirical issues in business cycles and other problems involving cyclical movements and phase leads. The reader is recommended to visit Estima's website where informative newsletters are published and useful procedures are logged and updated at some frequencies.

LIMDEP

http://www.limdep.com

LIMDEP, according to its name, was initially specialised in estimation of models involving limited dependent variables, models introduced in Chapter 11, Chapter 12 and to some extent, Chapter 13 of this book. However, having been developed and evolved over years, it is now a complete econometrics package, as claimed by its developer and provider, Econometric Software, Inc. As such, LIMDEP Version 9.0 is an integrated programme for estimation and analysis of linear and nonlinear models, with cross section, time series and panel data. Since LIMDEP has long been a leader in the fields of discrete choice, censoring and truncation, panel data analysis, frontier and efficiency estimation, its collection of procedures for analysing these models is very comprehensive compared with other packages. The main feature of the package is a suite of more than 100 built-in estimators for all forms of the linear regression model, and stochastic frontier, discrete choice and limited dependent variable models, including models for binary, censored, truncated, survival, count, discrete and continuous variables and a variety of sample selection models. LIMDEP is widely used for teaching and research by higher education institutions, government departments and agencies, research establishments, business communities and industries around the world.

TSP

http://www.tspintl.com

To some extent, TSP, or Time Series Processor, is similar to RATS. So we do not introduce it in detail and the reader can refer to the website of TSP International for detail.

GAUSS

http://www.aptech.com

GAUSS is powerful in matrix operations. The GAUSS Mathematical and Statistical System is a fast matrix programming language, one of the most popular software packages for economists and econometricians as well as for scientists, engineers, statisticians, biometricians, and financial analysts. Designed for computationally intensive tasks, the GAUSS system is ideally suited for the researcher who does not have the time required to develop programs in C or FORTRAN but finds that most statistical or mathematical 'packages' are not flexible or powerful enough to perform complicated analysis or to work on large problems. Compared with RATS, GAUSS is more power-ful and efficient but requires higher levels of programming knowledge and skills.

Microfit

http://www.econ.cam.ac.uk/microfit

Microfit is a menu-driven easy to use econometric package written especially for microcomputers, and is specifically designed for the econometric modelling of time series data. The strength of the package lies in the fact that it can be used at different levels of technical sophistication. For experienced users of econometric programmes it offers a variety of univariate and multivariate estimation methods and provides a large number of diagnostic and non-nested tests not readily available in other packages. As a result, Microfit is one of the most frequently used econometric packages by economists and applied econometricians.

SAS

http://www.sas.com

SAS is a large multi-purpose statistical package. It can process almost all model estimation problems in this book. But as it is large it is not usually available on PCs. It also requires more knowledge in software.

Matlab

http://www.mathworks.com

Matlab was initially developed for solving engineering problems. Now there are more and more economists and econometricians using this package. It is widely used for optimisation, control system design, signal and image processing and communications. Of the most relevance to this book are statistics and data analysis, and financial modelling and analysis.

Mathematica

http://www.wolfram.com/products/mathematica

Economists and econometricians increasingly use this package as well. Its mathematics and algorithms cover matrices, algebra, optimisation, statistics, calculus, discrete mathematics and number theory. With regard to statistics and data analysis, Mathematica provides integrated support both for classical statistics and for modern large-scale data analysis. Its symbolic character allows broader coverage, with symbolic manipulation of statistical distributions, symbolic specification of functions and models, and general symbolic representations of large-scale data properties. Incorporating the latest numerics and computational geometry algorithms, Mathematica provides high-accuracy and high-reliability statistical results for datasets of almost unlimited size.

14.3. Learned societies and professional associations

This section lists major learned societies and professional associations in the fields of finance, economics, econometrics as well as real estate and accounting.

American Finance Association (AFA)

http://www.afajof.org

The American Finance Association is the premier academic organisation devoted to the study and promotion of knowledge about financial economics. The AFA was planned at a meeting in December 1939 in Philadelphia. *The Journal of Finance* was first published in August 1946. Association membership has grown steadily over time and the AFA currently has over 8,000 members. The AFA sponsors an annual meeting in each January, usually at the same city and during the same days as the American Economic Association.

American Economic Association (AEA)

http://www.vanderbilt.edu/AEA

The American Economic Association was organised in 1885 at Saratoga, New York. Approximately 22,000 economists are members and there are 5,500 institution subscribers. Over 50 per cent of the membership is associated with academic institutions around the world, 35 per cent with business and industry and the remainder largely with US federal, state, and local government agencies. The Mission Statement of the AEA is: the encouragement of economic research, especially the historical and statistical study of the actual conditions of industrial life; the issue of publications on economic subjects; and the encouragement of perfect freedom of economic discussion, including an Annual Meeting (in each January). The Association as such will take no partisan attitude, nor will it commit its members to any position on practical economic questions. The three traditional publications by the AEA are amongst the most influential, with *American Economic Review* being one of the oldest, starting in 1911, *Journal of Economic Literature* being renowned for its classification system, and *Journal of Economic Perspectives* being unique in attempting to fill the gap between the general interest press and most other academic economics journals. The AEA has scheduled to launch four new journals in 2009: *Applied Economics*, *Economic Policy*, *Macroeconomics* and *Microeconomics* in an age of journal proliferation. The impact has yet to be appreciated.

American Accounting Association (AAA)

http://aaahq.org

The American Accounting Association promotes worldwide excellence in accounting education, research and practice. Founded in 1916 as the American Association of University Instructors in Accounting, its present name was adopted in 1936. The Association is a voluntary organisation of persons interested in accounting education and research. The mission of the AAA is to foster worldwide excellence in the creation, dissemination and application of accounting knowledge and skills.

The AAA publishes *The Accounting Review*, *Accounting Horizons* and *Issues in Accounting Education*.

Econometric Society

http://www.econometricsociety.org

The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics. The Econometric Society was founded in 1930, at the initiative of the Yale economist Irving Fisher (the Society's first president) and the Norwegian economist Ragnar Frisch, who some forty years later was the first economist (together with Jan Tinbergen) to be awarded the Nobel Prize. The first organisational meeting of the Society was held in Cleveland, Ohio, on 29 December 1930. The first scientific meetings of the Society were held in September 1931, at the University of Lausanne, Switzerland, and in December 1931, in Washington D.C.

   The journal *Econometrica* published its first issue in 1933, with Frisch as editor-in-chief, and with a budget that was initially subsidised by the financier Alfred Cowles. Frisch had coined the word 'econometrics' only a few years earlier in 1926. The journal started out publishing four issues of 112 pages per year and did not grow beyond 500 pages per year until the 1950s. Since the 1970s *Econometrica* has published six issues per year containing roughly 1,600 annual pages.

Financial Management Association International (FMA)

http://www.fma.org

Established in 1970, the Financial Management Association International is a global leader in developing and disseminating knowledge about financial decision making. The mission of the FMA is to broaden the common interests between academicians and practitioners, provide opportunities for professional interaction between and among academicians, practitioners and students, promote the development and understanding of basic and applied research and of sound financial practices, and to enhance the quality of education in finance. FMA's members include finance practitioners and academicians and students who are interested in the techniques and advances which define the field of finance. Over 5,000 academicians and practitioners throughout the world are members of the FMA. The FMA publishes *Financial Management*, *Journal of Applied Finance* and *FMA Survey and Synthesis Series*.

European Economic Association (EEA)

http://www.eeassoc.org

The European Economic Association, launched in 1984, is an international scientific body, with membership open to all persons involved or interested in

economics. The aims of the EEA are: to contribute to the development and application of economics as a science in Europe; to improve communication and exchange between teachers, researchers and students in economics in the different European countries; and to develop and sponsor co-operation between teaching institutions of university level and research institutions in Europe. In pursuing these aims, the EEA is particularly eager to foster closer links between theory-oriented and policy-oriented economists, as well as between students and more senior economists, from all parts of Europe. The EEA holds annual congresses, usually in August, and summer schools. The primary publication of the EEA is *The Journal of the European Economic Association*, known as *European Economic Review* prior to 2003. It is, though, a little eerie with such a title change, at a time when the attachment of the readership and authorship of a journal to its sponsoring association becomes decreasingly relevant. Other major scholarly associations have just gone through a changeover process in harmony with the trend. For example, we will see in the following that *European Finance Review*, of the European Finance Association, was renamed *Review of Finance* in 2004, and *The American Real Estate and Urban Economics Association Journal* was renamed as *Real Estate Economics* in 1995. The EEA also publishes *Economic Policy*.

European Finance Association (EFA)

http://www.efa-online.org

The European Finance Association was established in 1974 and is the oldest finance association in Europe. *European Finance Review*, renamed *Review of Finance* in 2004, is the association's main publication and is regarded as one of leading finance journals in the world. The association holds annual meetings all over Europe, usually in August. Participation of the annual meeting is not confined to Europe and geography does not play a role in any major aspects of the annual meeting, including topics and organisation.

European Financial Management Association (EFMA)

http://www.efmaefm.org

The European Financial Management Association was founded in 1994 to encourage research and disseminate knowledge about financial decision making in all areas of finance as it relates to European corporations, financial institutions and capital markets. 20 years after the establishment of the EFA, the launch of the EFMA has intensified finance research activity in Europe to a great extent through introducing competition. The Association's membership consists of academics, practitioners and students from Europe and the rest of the world who are interested in the practice of sound financial management techniques and are dedicated in understanding and solving financial problems. The EFMA holds annual meetings in each June and publishes *European Financial Management*.

Royal Economic Society (RES)

http://www.res.org.uk

Established in the nineteenth century, the Royal Economic Society is one of the oldest economic associations in the world. Currently it has over 3,000 individual members, of whom 60 per cent live outside the United Kingdom. It is a professional association which promotes the encouragement of the study of economic science in academic life, government service, banking, industry and public affairs. *The Economic Journal*, first published the 1990s, is one of the oldest in the world. The RES launched a new journal *The Econometrics Journal* in 1998.

American Real Estate and Urban Economics Association (AREUEA)

http://www.areuea.org

The American Real Estate and Urban Economics Association was originated at the 1964 meeting of the Allied Social Science Association in Chicago. The AREUEA grew from discussions of individuals that recognised a need for more information and analysis in the fields of real estate development, planning and economics. The continuing efforts of this non-profit association has advanced the scope of knowledge in these disciplines and has facilitated the exchange of information and opinions among academic, professional and governmental people who are concerned with urban economics and real estate issues. The AREUEA's journal, *Real Estate Economics* (formerly *The American Real Estate and Urban Economics Association Journal* prior to 1995) is published quarterly and is distributed on a calendar year subscription basis. The journal contains research and scholarly studies of current and emerging real estate issues.

American Real Estate Society (ARES)

http://www.aresnet.org

American Real Estate Society was founded in 1985 to serve the educational, informational, and research needs of thought leaders in the real estate industry and real estate professors at colleges and universities. The REAS has several affiliated societies, with the largest, The International Real Estate Society (IRES), being founded in 1993. The REAS publishes *Journal of Real Estate Research*, *Journal of Real Estate Literature*, and *Journal of Real Estate Portfolio Management*.

International Institute of Forecasters (IIF)

http://www.forecasters.org

The International Institute of Forecasters' objectives are to stimulate the generation, distribution, and use of knowledge on forecasting. The IIF was founded in 1981 as a non-profit organisation. The IIF sponsors an annual International Symposium of Forecasting in each June and publishes *International Journal of Forecasting*.

14.4.   Organisations and institutions

14.4.1.   *International financial institutions and other organisations*

International Monetary Fund (IMF)

http://www.imf.org

The IMF is an international organisation of 185 member countries, established to promote international monetary co-operation, exchange stability, and orderly exchange arrangements; to foster economic growth and high levels of employment; and to provide temporary financial assistance to countries to help ease balance of payments adjustment. Since the IMF was established in 1946, its purposes have remained unchanged but its operations, which involve surveillance, financial assistance, and technical assistance, have developed to meet the changing needs of its member countries in an evolving world economy.

World Bank

http://www.worldbank.org

The World Bank is the world's largest financial source of development assistance. It consists of two development institutions owned by 185 member countries, the International Bank for Reconstruction and Development (IBRD) and the International Development Association (IDA). Since its inception in 1944, the Bank uses its financial resources, highly trained staff, and extensive knowledge base to help each developing country onto a path of stable, sustainable, and equitable growth in the fight against poverty.

Organisation for Economic Co-operation and Development (OECD)

http://www.oecd.org

The Organisation for Economic Co-operation and Development has been called a think tank, and monitoring agency. Evolved from the Organisation for European Economic Co-operation (OEEC), the OECD was formed in 1961. The OECD groups 30 member countries in an organisation that, most importantly, provides governments a setting in which to discuss, develop and perfect economic and social policy. It is rich, in that OECD countries produce two thirds of the world's goods and services, but it is not an exclusive club. Essentially, membership is limited only by a country's commitment to a market economy and a pluralistic democracy. The core of original members has expanded from Europe and North America to include Japan, Australia, New Zealand, Finland, Mexico, the Czech Republic, Hungary, Poland and Korea. There are many more contacts with the rest of the world through programmes with countries in the former Soviet bloc, Asia, and Latin America, which, in some cases, may lead to membership.

Exchanges between OECD governments are facilitated by information and analysis provided by a Secretariat in Paris. Parts of the OECD Secretariat collect

data, monitor trends, analyse and forecast economic developments, while others research social changes or evolving patterns in trade, environment, agriculture, technology, taxation and more. This work, in areas that mirror the policy-making structures in ministries of governments, is done in close consultation with policy-makers who will use the analysis, and it underpins discussion by member countries when they meet in specialised committees of the OECD. Much of the research and analysis is published.

European Bank for Reconstruction and Development (EBRD)

http://www.ebrd.org

The European Bank for Reconstruction and Development was established in 1991. It exists to foster the transition towards open market-oriented economies and to promote private and entrepreneurial initiative in the countries of central and eastern Europe and the Commonwealth of Independent States (CIS) committed to and applying the principles of multiparty democracy, pluralism and market economics.

Asian Development Bank (ADB)

http://www.adb.org

The ADB is a multilateral development finance institution dedicated to reducing poverty in Asia and the Pacific. Established in 1966, the ADB is now owned and financed by 67 member countries, of which 48 countries are from the region and the rest from the other parts of the world. The ADB helps improve the quality of people's lives by providing loans and technical assistance for a broad range of development activities.

Bank for International Settlements (BIS)

http://www.bis.org

This website has links to all central banks websites. The BIS is an international organisation which fosters co-operation among central banks and other agencies in pursuit of monetary and financial stability. The BIS functions as: a forum for international monetary and financial co-operation where central bankers and others meet and where facilities are provided to support various committees, both standing and *ad hoc;* a bank for central banks, providing a broad range of financial services; a centre for monetary and economic research, contributing to a better understanding of international financial markets and the interaction of national monetary and financial policies; an agent or trustee, facilitating the implementation of various international financial agreements. The BIS operates the Financial Stability Institute (FSI) jointly with the Basel Committee on Banking Supervision. The BIS also hosts the secretariats of the Financial Stability Forum (FSF) and the International Association of Insurance Supervisors (IAIS).

14.4.2. Major stock exchanges, option and futures exchanges, and regulators

New York Stock Exchange (NYSE)

http://www.nyse.com

The New York Stock Exchange traces its origins to a founding agreement, the Buttonwood Agreement by 24 New York City stockbrokers and merchants, in 1792. The NYSE registered as a national securities exchange with the US Securities and Exchange Commission on 1 October 1934. The Governing Committee was the primary governing body until 1938, at which time the Exchange hired its first paid president and created a thirty-three member Board of Governors. The Board included Exchange members, non-member partners from both New York and out-of-town firms, as well as public representatives.

In 1971 the Exchange was incorporated as a not-for-profit corporation. In 1972 the members voted to replace the Board of Governors with a twenty-five member Board of Directors, comprised of a Chairman and CEO, twelve representatives of the public, and twelve representatives from the securities industry. Subject to the approval of the Board, the Chairman may appoint a President, who would serve as a director. Additionally, at the Board's discretion, they may elect an Executive Vice Chairman, who would also serve as a director. On 4 April 2007, NYSE Euronext was created by the combination of NYSE Group and Euronext (to be introduced later), which brings together six cash equities exchanges in five countries of the US, France, the Netherlands, Belgium and Portugal, and six derivatives exchanges in these five countries plus the UK, where London International Financial Futures and Options Exchange operates, which became part of Euronext in 2002. Prior to the merger with Euronext, the NYSE merged with the Archipelago Exchange (ArcaEx) in 2006 and the latter acquired the Pacific Exchange a year earlier.

London Stock Exchange (LSE)

http://www.londonstockexchange.com

The London Stock Exchange was formed in 1760 by 150 brokers as a club for share trading. It changed to the current name in 1773. Since 1986, trading has moved from being conducted face-to-face on a market floor to being performed via computer and telephone from separate dealing rooms. This is due to the introduction of SEAQ and SEAQ International, two computer systems displaying share price information in brokers' offices around the UK. The LSE became a private limited company under the Companies Act 1985. In 1991 the Exchange replaced the governing Council of the Exchange with a Board of directors drawn from the Exchange's executive, customer and user base. In recent years there have been major changes in the Exchange: the role of the Exchange as UK listing authority with the Treasury was transferred to the Financial Services Authority in 2000; and the Exchange became a Plc and has been listed since July 2001.

Tokyo Stock Exchange (TSE)

http://www.tse.or.jp

In the 1870s, a securities system was introduced in Japan and public bond negotiation began. This resulted in the request for a public trading institution and the 'Stock Exchange Ordinance' was enacted in May 1878. Based on this ordinance, the 'Tokyo Stock Exchange Co. Ltd.' was established on 15 May 1878 and trading began on 1 June. The TSE functions as a self-regulated, non-profit association. Established under a provision of the Securities and Exchange Law, the TSE is managed and maintained by its members.

National Association of Securities Dealers Automated Quotations (NASDAQ)

http://www.nasdaq.com

The National Association of Securities Dealers Automated Quotations is the first and the largest electronic screen based stock trading market. The NASDAQ was established on 8 February 1971. With approximately 3,200 companies, it lists more companies and, on average, trades more shares per day than any other market. As a response to changes in technology and a challenge to the traditional trading system, the NASDAQ was referred to as an over-the-counter trading arrangement as late as in 1987. However, its growth and development seem unstoppable in an era of global technology advances. In November 2007, the NASDAQ acquired the Philadelphia Stock Exchange, the oldest stock exchange in the US.

Chicago Mercantile Exchange (CME)

http://www.cme.com

The Chicago Mercantile Exchange was established in 1919, evolved from the Chicago Butter and Egg Board founded in 1898. Initially, its members traded futures contracts on agricultural commodities via open outcry. This system of trading, which is still in use today, essentially involves hundreds of auctions going on at the same time. Open outcry is an efficient means of 'price discovery', a term widely referred to one of the roles played by futures markets. Its speed and efficiency have been further enhanced by the introduction of a variety of trading floor technologies. Nowadays, the CME open outcry platform and trading floor systems are linked to the CME® Globex® electronic trading platform, which allows market participants to buy and sell the products almost wherever and whenever, at trading booths on its Chicago trading floors, at offices or homes, during and after regular trading hours.

In 2007, the new CME group was formed by the CME's merger with the Chicago Board of Trade (to be introduced next), the latter being the oldest derivatives exchange in the world. Nevertheless, the two derivatives exchanges had traded each other's products long before the merger. The CME group is the largest and

most diverse derivatives exchange in the world. The combined volume of trades in 2006 exceeded 2.2 billion contracts, worth more than $1,000 trillion. Three quarters of its trades are executed electronically.

Chicago Board of Trade (CBOT)

http://www.cbot.com

The Chicago Board of Trade was established in 1848 and is the world's oldest derivatives exchange. More than 3,600 CBOT members trade 48 different futures and options products at the CBOT, resulting in 2000 annual trading volume of more than 233 million contracts. Early in its history the CBOT listed for trading only agricultural instruments, such as wheat, corn and oats. In 1975, the CBOT expanded its offering to include financial contracts, initially, the US Treasury Bond futures contract which is now one of the world's most actively traded.

The CBOT presently is a self-governing, self-regulated not-for-profit, non-stock corporation that serves individuals and member firms. The governing body of the exchange consists of a Board of Directors that includes a Chairman, First Vice Chairman, Second Vice Chairman, 18 member directors, five public directors, and the President. The Exchange is administered by an executive staff headed by the President and Chief Executive Officer. In 2007, the CBOT and the CME have merged to form the new CME group, reclaiming the CBOT's position as one of the leading and dominant derivatives exchanges in the world.

Chicago Board Options Exchange (CBOE)

http://www.cboe.com

The Chicago Board Options Exchange was founded in 1973. Prior to that time, options were traded on an unregulated basis and did not have to adhere to the principle of 'fair and orderly markets'. At the opening on 26 April 1973, the CBOE traded call options on 16 underlying stocks. Put options were introduced in 1977. By 1975, options had become so popular that other securities exchanges began entering the business. The quick acceptance of listed options propelled CBOE to become the second largest securities exchange in the country and the world's largest options exchange. In 1983, options on stock indices were introduced by the CBOE. Today, the CBOE accounts for more than 51 per cent of all US options trading and 91 per cent of all index options trading. The CBOE now lists options on over 1,200 widely traded stocks.

The CBOE was originally created by the CBOT but has always been managed and regulated as an independent entity. Due to increased volume in the early 1980s, the CBOE outgrew its trading facilities at the CBOT and moved into its own building in 1984.

Euronext

http://www.euronext.com

Euronext is a pan-European financial market, consisting of the Amsterdam Exchange, the Brussels Exchange, the Lisbon Exchange, the London International Financial Futures and Options Exchange (to be introduced next), and the Paris Bourse. It was formed in 2000 through a merger between the Paris, Amsterdam and Brussels exchanges and was listed on 6 June the next year. In 2002, the Lisbon Exchange and the London's derivatives exchange become part of Euronext. On 4 April 2007, NYSE Euronext was created by the merger between the NYSE Group and Euronext.

The history of Euronext can be traced back to 1602 when the Amsterdam Stock Exchange started to take shape. At the time, the Dutch East India Company, the world's first company to issue shares of stock on a large scale, was established, fostering the Amsterdam Stock Exchange. The Paris Bourse was created in 1724, several decades ahead of the establishment of the NYSE and the LSE. The Lisbon Exchange, then known as the Business Man's Assembly, was established in 1769; and the Brussels Stock Exchange was created in 1801.

London International Financial Futures and Options Exchange (LIFFE)

http://www.liffe.com

Unlike the case of the CBOT and CBOE where the latter was created by the former and they are now separated exchanges, the LIFFE was created when the original LIFFE, the London International Financial Futures Exchange, merged with the London options exchange. Notice that the acronym does not include the first letter of Options though Options is with the full name. In February 1999, the LIFFE's shareholders voted unanimously for a corporate restructuring which progressed the LIFFE further towards becoming a profit-oriented commercial organisation. With effect from April 1999, the restructuring split the right to trade and membership from shareholding, simplified a complex share structure and enabled non-members to purchase shares in LIFFE (Holdings) plc. In 2002, the LIFFE was acquired by Euronext and the latter merged with the NYSE in 2007. So, the LIFFE is now part of the NYSE Euronext group.

Philadelphia Stock Exchange (PHLX)

http://www.phlx.com

The Philadelphia Stock Exchange was founded in 1790 as the first organised stock exchange in the United States. The PHLX trades more than 2,200 stocks, 922 equity options, 10 index options, and 100 currency pairs.

The PHLX is reputed for its invention of exchange traded currency options in 1982. By 1988, currency options were trading in volumes as high as $4 billion per day in underlying value. Currency options put the Exchange on international maps, bringing trading interest from Europe, Pacific Rim and the Far East, and leading the

Exchange to be the first securities exchange to open international offices in money centres overseas. Currency options made the PHLX an around-the-clock operation. In September 1987, Philadelphia was the first securities exchange in the United States to introduce an evening trading session, chiefly to accommodate increasing demand for currency options in the Far East, and the exchange responded to growing European demand by adding an early morning session in January 1989. In September 1990, The PHLX became the first exchange in the world to offer around-the-clock trading by bridging the gap between the night session and the early morning hours. Although the exchange subsequently scaled back its trading hours, its current currency option trading hours from 2:30 a.m. to 2:30 p.m. (the Philadelphia time) are longer than any other open outcry auction marketplace. The PHLX was acquired by the NASDAQ in November 2007.

Shanghai Stock Exchange (SSE)

http://www.sse.com.cn

The Shanghai Stock Exchange was founded on 26 November 1990 and started trading on 19 December the same year. It is a non-profit-making membership institution regulated by the China Securities Regulatory Commission. Located in Shanghai, the SEE has enjoyed its geographical advantages, not only because of Shanghai's status as the financial centre of the PRC, but also because of its neighbouring provinces' technology and manufacturing muscles. There were over 37.87 million investors and 837 listed companies by the end of December 2004. The total market capitalisation of SSE listed companies reached RMB 2.6 trillion and capital raised from the SSE exceeded RMB 45.7 billion in 2004.

Shenzhen Stock Exchange (SZSE)

http://www.szse.cn

The Shenzhen Stock Exchange was established on 1 December 1990. Like the SSE, the SZSE is regulated by the China Securities Regulatory Commission. Shenzhen is located in southern Guangdong province, opposite Hong Kong. Chosen to be one of the special economic zones, the city of Shenzhen has developed rapidly in the last three decades and it appeared logical for the city of Shenzhen, along with Shanghai, to house one of the two stock exchanges. There were 579 companies listed on the SZSE, the total market capitalisation of these companies amounted to RMB 1.8 trillion and RMB 62.1 billion of funds was raised on the SZSE in 2006.

Securities and Exchange Commission (SEC)

http://www.sec.gov

The SEC's foundation was laid in an era that was ripe for reform. Before the Great Crash of 1929, there was little support for federal regulation of the securities markets. This was particularly true during the post-World War I surge of securities activity. Proposals that the federal government require financial disclosure and

prevent the fraudulent sale of stock were never seriously pursued. Tempted by promises of 'rags to riches' transformations and easy credit, most investors gave little thought to the dangers inherent in uncontrolled market operation. During the 1920s, approximately 20 million large and small shareholders took advantage of post-war prosperity and set out to make their fortunes in the stock market. It is estimated that of the $50 billion in new securities offered during this period, half became worthless.

The primary mission of the US Securities and Exchange Commission (SEC) is to protect investors and maintain the integrity of the securities markets. As more and more first-time investors turn to the markets to help secure their futures, pay for homes, and send children to college, these goals are more compelling than ever.

The laws and rules that govern the securities industry in the United States derive from a simple and straightforward concept: all investors, whether large institutions or private individuals, should have access to certain basic facts about an investment prior to buying it. To achieve this, the SEC requires public companies to disclose meaningful financial and other information to the public, which provides a common pool of knowledge for all investors to use to judge for themselves if a company's securities are a good investment. Only through the steady flow of timely, comprehensive and accurate information can people make sound investment decisions. The SEC also oversees other key participants in the securities world, including stock exchanges, broker-dealers, investment advisors, mutual funds, and public utility holding companies. Here again, the SEC is concerned primarily with promoting disclosure of important information, enforcing the securities laws, and protecting investors who interact with these various organisations and individuals.

Crucial to the SEC's effectiveness is its enforcement authority. Each year the SEC brings between 400–500 civil enforcement actions against individuals and companies that break the securities laws. Typical infractions include insider trading, accounting fraud, and providing false or misleading information about securities and the companies that issue them. Fighting securities fraud, however, requires teamwork. At the heart of effective investor protection is an educated and careful investor. The SEC offers the public a wealth of educational information on its Internet website at www.sec.gov. The website also includes the EDGAR database of disclosure documents that public companies are required to file with the Commission.

Though it is the primary overseer and regulator of the US securities markets, the SEC works closely with many other institutions, including Congress, other federal departments and agencies, the self-regulatory organisations, e.g. stock exchanges, state securities regulators, and various private sector organisations.

China Securities Regulatory Commission (CSRC)

http://www.csrc.gov.cn

The China Securities Regulatory Commission is a ministry level agency of the State Council (central administration) of the PRC, established in 1992. It is modelled

after the US SEC and functions similarly to that of the SEC. Its basic functions are: to establish a centralised supervisory system for securities and futures markets and to assume direct leadership over securities and futures market supervisory bodies; to strengthen the supervision over securities and futures and related business, including stock and futures exchange markets, the listed companies, and companies engaged in securities trading, investment and consultancy, and raise the standard of information disclosure; to increase the abilities to prevent and handle financial crisis; to organise the drafting of laws and regulations for securities markets, study and formulate the principles, policies and rules related to securities markets, formulate development plans and annual plans for securities markets, and direct, co-ordinate, supervise and examine matters related to securities in various regions and relevant departments.

Financial Services Authority (FSA)

www.fsa.gov.uk

The FSA is a relatively new organisation, which was founded in 1997 when the UK government announced its decision to merge the supervision of banking and investment under one regulatory organisation. This created a tripartite regulatory system and arrangements involving the Treasury, the Bank of England and the FSA. The FSA's predecessor was known as the Securities and Investments Board (SIB), created in May 1997 and changed to its current name in October 1997. The FSA is an independent, non-governmental body. The FSA board is appointed by an executive chairman, with three managing directors and eleven non-executive directors.

The first stage of the recent reform of UK financial services regulation was completed in June 1998, when responsibility for banking supervision was transferred to the FSA from the Bank of England. The Financial Services and Markets Act of the UK, which received Royal Assent in June 2000 and was implemented on 1 December 2001, transferred to the FSA the responsibilities of several other organisations: Building Societies Commission, Friendly Societies Commission, Investment Management Regulatory Organisation, Personal Investment Authority, Register of Friendly Societies, and Securities and Futures Authority.

The FSA regulates the financial services industry and has four objectives under the Financial Services and Markets Act 2000: maintaining market confidence; promoting public understanding of the financial system; the protection of consumers; and fighting financial crime. The FSA is the UK's financial watchdog, keeping an eye on the goings-on in the City. It regulates and oversees the financial system, and plays an important part in ensuring that training within the banking industry is up to scratch. Find out more about the FSA's important role in this briefing. All companies in the financial services market, from banks to pension companies, must be FSA accredited. After accreditation, these companies are supervised and inspected on a regular basis. The FSA imposes levies on accredited companies.

Following the Northern Rock fiasco, the FSA has experienced some shake-ups in its supervision arrangements and functions, one of which is to increase resources for its overall banking sector capability. This has enhanced the FSA's power in the UK's tripartite regulatory system and rationalised the coordination between the three parties to a certain extent.

China Banking Regulatory Commission (CBRC)

http://www.cbrc.gov.cn

The China Banking Regulatory Commission and the China Insurance Regulatory Commission (to be introduced next) are ministry level agencies of the State Council of the PRC, with the former focusing on the supervision of banking institutions and the latter on the supervision of insurance companies. The CBRC was established in 2003, with its functions being partly transferred from the People's Bank of China, the central bank, and partly newly created due to the change and expansion in the banking sector. The CBRC, together with the China Insurance Regulatory Commission, is modelled after the FSA of the UK to a certain extent, the wisdom of which has yet to be tested. The main functions of the CBRC include: to formulate supervisory rules and regulations governing banking institutions; to authorise the establishment, changes, termination and business scope of banking institutions; to conduct on-site examination and off-site surveillance of banking institutions, and take enforcement actions against rule-breaking behaviour; and to conduct fit-and-proper tests on the senior managerial personnel of banking institutions; to compile and publish statistics and reports of the overall banking industry in accordance with relevant regulations.

China Insurance Regulatory Commission (CIRC)

http://www.circ.gov.cn

The China Insurance Regulatory Commission is a ministry level agency of the State Council of the PRC, a sister agency of the CBRC. Established in 1998 as a sub-ministry level agency, the CIRC was upgraded to the ministry level in 2003. Its main functions cover: To formulate guidelines and policies for developing insurance business, draw up development strategies and plans for the insurance industry; to formulate laws, rules and regulations for insurance supervision, and rules and regulations for the industry; to approve the establishment of insurance companies in various forms; to approve the categories of insurance schemes related to public interests, impose insurance articles and rates of premium for compulsory insurance schemes and newly developed life insurance schemes, and accept filing for the record of articles and premium rates of other insurance schemes; to supervise the payment ability and market conduct of insurance companies; to supervise policy-oriented insurance and compulsory insurance operations; and to investigate and mete out punishment against unfair competition and illegal conduct of insurance institutions and individuals as well as the operations of non-insurance institutions and disguised insurance operations.

14.4.3. Central banks

Board of Governors of the Federal Reserve System

http://www.federalreserve.gov

The Federal Reserve, the central bank of the United States, was founded by Congress in 1913 to provide the nation with a safer, more flexible, and more stable monetary and financial system. Today the Federal Reserve's duties fall into four general areas: conducting the nation's monetary policy; supervising and regulating banking institutions and protecting the credit rights of consumers; maintaining the stability of the financial system; and providing certain financial services to the US government, the public, financial institutions, and foreign official institutions.

The Federal Reserve System was designed to ensure its political independence and its sensitivity to divergent economic concerns. The chairman and the six other members of the Board of Governors who oversee the Federal Reserve are nominated by the President of the United States and confirmed by the Senate. The President is directed by law to select governors who provide 'a fair representation of the financial, agricultural, industrial and geographical divisions of the country'.

Each Reserve Bank is headed by a president appointed by the Bank's nine-member board of directors. Three of the directors represent the commercial banks in the Bank's region that are members of the Federal Reserve System. The other directors are selected to represent the public with due consideration to the interest of agriculture, commerce, industry, services, labor and consumers. Three of these six directors are elected by member banks and the other three are chosen by the Board of Governors.

The 12 Reserve Banks supervise and regulate bank holding companies as well as state chartered banks in their District that are members of the Federal Reserve System. Each Reserve Bank provides services to depository institutions in its respective District and functions as a fiscal agent of the US government.

Federal Reserve Bank of New York

http://www.ny.frb.org

The Federal Reserve Bank of New York is one of 12 regional Reserve Banks which, together with the Board of Governors in Washington, D.C., comprise the Federal Reserve System. Stored inside the vaults of the New York Fed building is hundreds of billions of dollars of gold and securities. But what is unique and most significant about the Bank is its broad policy responsibilities and the effects of its operations on the US economy.

The New York Fed has supervisory jurisdiction over the Second Federal Reserve District, which encompasses New York State, the 12 northern counties of New Jersey, Fairfield County in Connecticut, Puerto Rico and the Virgin Islands. Though it serves a geographically small area compared with those of other Federal

Reserve Banks, the New York Fed is the largest Reserve Bank in terms of assets and volume of activity.

European Central Bank (ECB)

http://www.ecb.int

The European System of Central Banks (ESCB) is composed of the European Central Bank (ECB) and the national central banks (NCBs) of all 15 EU Member States. The 'Eurosystem' is the term used to refer to the ECB and the NCBs of the Member States which have adopted the euro. The NCBs of the Member States which do not participate in the euro area, however, are members of the ESCB with a special status – while they are allowed to conduct their respective national monetary policies, they do not take part in the decision-making with regard to the single monetary policy for the euro area and the implementation of such decisions.

In accordance with the Treaty establishing the European Community (the 'Treaty') and the Statute of the European System of Central Banks and of the European Central Bank (the 'Statute'), the primary objective of the Eurosystem is to maintain price stability. Without prejudice to this objective, it shall support the general economic policies in the Community and act in accordance with the principles of an open market economy.

The basic tasks to be carried out by the Eurosystem are: to define and implement the monetary policy of the euro area; to conduct foreign exchange operations; to hold and manage the official foreign reserves of the Member States; and to promote the smooth operation of payment systems.

People's Bank of China

http://www.pbc.gov.cn

The central bank of the People's Republic of China is the People's Bank of China. In the early times of the People's Republic, the Bank, though a ministerial department of the State Council, was co-ordinated by the Ministry of Finance. It was largely a commercial bank with high street branches all over the country. During the 1980s, its commercial and corporate banking functions were reorganised and grouped into a new and separate bank, the Industrial and Commercial Bank of China, probably the largest bank in the world. Since then the People's Bank plays solely the role of a central bank and is completely independent of the Ministry of Finance but is not separated from the Administration. The People's Bank used to have branches at the province's level for regional monetary policy matters or the monitoring of monetary policy of the Headquarters, which have now been reorganised to form regional branches (each covering several provinces/municipal cities), a structure similar to that of the US Federal Reserve System. Following the establishment of the China Banking Regulatory Commission in 2003, the role and responsibilities for banking

supervision of the People's Bank have been transferred to the newly founded commission. The People's Bank now plays an important role solely in the areas of monetary policy.

Bank of Japan

http://www.boj.or.jp

The role of the Bank of Japan is similar to that of the pre-1997 Bank of England in that the Treasury or the Ministry of Finance makes important decisions and monetary policy and the Bank implements monetary policy.

The Bank of Japan's missions are to maintain price stability and to ensure the stability of the financial system, thereby laying the foundations for sound economic development. To fulfil these two missions, the Bank conducts the following activities: issuance and management of banknotes; implementation of monetary policy; providing settlement services and ensuring the stability of the financial system; treasury and government securities-related operations; international activities; and compilation of data, economic analyses and research activities.

Bank of Russia

http://www.cbr.ru

The Central Bank of the Russian Federation (Bank of Russia) was founded on 13 July 1990, on the basis of the Russian Republic Bank of the State Bank of the USSR. Accountable to the Supreme Soviet of the RSFSR, it was originally called the State Bank of the RSFSR. In November 1991 when the Commonwealth of Independent States was founded and Union structures dissolved, the Supreme Soviet of the RSFSR declared the Central Bank of the RSFSR to be the only body of state monetary and foreign exchange regulation in the RSFSR. The functions of the State Bank of the USSR in issuing money and setting the rouble exchange rate were transferred to it. The Central Bank of the RSFSR was instructed to assume, before 1 January 1992, full control of the assets, technical facilities and other resources of the State Bank of the USSR and all its institutions, enterprises and organisations.

The Bank of Russia carries out its functions, which were established by the Constitution of the Russian Federation (Article 75) and the Law on the Central Bank of the Russian Federation (Bank of Russia) (Article 22), independently from the federal, regional and local government structures. In 1992–1995, to maintain the stability of the banking system, the Bank of Russia set up a system of supervision and inspection of commercial banks and a system of foreign exchange regulation and foreign exchange control. As an agent of the Ministry of Finance, it organised a government securities market, known as the GKO market, and began to participate in its operations.

Schweizerische Nationalbank

http://www.snb.ch

Schweizerische Nationalbank, or the Swiss National Bank in English, is the central bank of Switzerland and, as such, conducts the country's monetary policy. It is obliged by the Constitution and by statute to act in accordance with the interests of the country as a whole. Its primary goal is to ensure price stability, while taking due account of economic developments, creating and promoting an appropriate environment for economic growth.

The Swiss National Bank has two head offices, one in Berne and one in Zurich, with a branch with cash distribution services in Geneva and five representative offices in major Swiss cities. Furthermore, it has 16 agencies operated by cantonal banks that help secure the supply of money to the country. The Bank Council oversees and controls the conduct of business by the Swiss National Bank while the Governing Board runs the bank.

Bank of Canada

http://www.bankofcanada.ca

The Bank of Canada is the central bank of the country. It has responsibilities for Canada's monetary policy, bank notes, financial system and funds management. Its principal role, as defined in the Bank of Canada Act, is to promote the economic and financial welfare of Canada. The Bank of Canada was founded in 1934 as a privately owned corporation. In 1938, it became a Crown corporation belonging to the federal government. Since that time, the Minister of Finance has held the entire share capital issued by the Bank. Ultimately, the Bank is owned by the people of Canada.

As the central bank, the Bank of Canada's four main areas of responsibility are as follows: to conduct monetary policy, with the goal being to contribute to solid economic performance and rising living standards for Canadians by keeping inflation low, stable, and predictable; to issue Canada's bank notes and be responsible for their design and security, distribution, and replacement; to actively promote safe, sound and efficient financial systems, both within Canada and internationally; and to provide high quality, effective, and efficient funds-management services for the Canadian federal government, the Bank of Canada, and other clients.

Index

Page references followed by e indicate a boxed example; f indicates an illustrative figure; t indicates a table