



# Introduction

This chapter sets the scene for the book by discussing in broad terms the questions of what is econometrics, and what are the ‘stylised facts’ describing financial data that researchers in this area typically try to capture in their models. It also collects together a number of preliminary issues relating to the construction of econometric models in finance.

---

## Learning Outcomes

In this chapter, you will learn how to

- Distinguish between different types of data
  - Describe the steps involved in building an econometric model
  - Calculate asset price returns
  - Construct a workfile, import data and accomplish simple tasks in EViews
- 

## 1.1 What is econometrics?

The literal meaning of the word econometrics is ‘measurement in economics’. The first four letters of the word suggest correctly that the origins of econometrics are rooted in economics. However, the main techniques employed for studying economic problems are of equal importance in financial applications. As the term is used in this book, financial econometrics will be defined as the *application of statistical techniques to problems in finance*. Financial econometrics can be useful for testing theories in finance, determining asset prices or returns, testing hypotheses concerning the relationships between variables, examining the effect on financial markets of changes in economic conditions, forecasting future values of financial variables and for financial decision-making. A list of possible examples of where econometrics may be useful is given in box 1.1.

**Box 1.1** The value of econometrics

- (1) Testing whether financial markets are weak-form informationally efficient
- (2) Testing whether the Capital Asset Pricing Model (CAPM) or Arbitrage Pricing Theory (APT) represent superior models for the determination of returns on risky assets
- (3) Measuring and forecasting the volatility of bond returns
- (4) Explaining the determinants of bond credit ratings used by the ratings agencies
- (5) Modelling long-term relationships between prices and exchange rates
- (6) Determining the optimal hedge ratio for a spot position in oil
- (7) Testing technical trading rules to determine which makes the most money
- (8) Testing the hypothesis that earnings or dividend announcements have no effect on stock prices
- (9) Testing whether spot or futures markets react more rapidly to news
- (10) Forecasting the correlation between the stock indices of two countries.

The list in box 1.1 is of course by no means exhaustive, but it hopefully gives some flavour of the usefulness of econometric tools in terms of their financial applicability.

## 1.2 Is financial econometrics different from ‘economic econometrics’?

As previously stated, the tools commonly used in financial applications are fundamentally the same as those used in economic applications, although the emphasis and the sets of problems that are likely to be encountered when analysing the two sets of data are somewhat different. Financial data often differ from macroeconomic data in terms of their frequency, accuracy, seasonality and other properties.

In economics, a serious problem is often a *lack of data at hand* for testing the theory or hypothesis of interest – this is often called a ‘small samples problem’. It might be, for example, that data are required on government budget deficits, or population figures, which are measured only on an annual basis. If the methods used to measure these quantities changed a quarter of a century ago, then only at most twenty-five of these annual observations are usefully available.

Two other problems that are often encountered in conducting applied econometric work in the arena of economics are those of *measurement error* and *data revisions*. These difficulties are simply that the data may be estimated, or measured with error, and will often be subject to several vintages of subsequent revisions. For example, a researcher may estimate an economic model of the effect on national output of investment in computer technology using a set of published data, only to find that the

data for the last two years have been revised substantially in the next, updated publication.

These issues are rarely of concern in finance. Financial data come in many shapes and forms, but in general the prices and other entities that are recorded are those at which trades *actually took place*, or which were *quoted* on the screens of information providers. There exists, of course, the possibility for typos and possibility for the data measurement method to change (for example, owing to stock index re-balancing or re-basing). But in general the measurement error and revisions problems are far less serious in the financial context.

Similarly, some sets of financial data are observed at much *higher frequencies* than macroeconomic data. Asset prices or yields are often available at daily, hourly, or minute-by-minute frequencies. Thus the number of observations available for analysis can potentially be very large – perhaps thousands or even millions, making financial data the envy of macro-econometricians! The implication is that more powerful techniques can often be applied to financial than economic data, and that researchers may also have more confidence in the results.

Furthermore, the analysis of financial data also brings with it a number of new problems. While the difficulties associated with handling and processing such a large amount of data are not usually an issue given recent and continuing advances in computer power, financial data often have a number of additional characteristics. For example, financial data are often considered very ‘noisy’, which means that it is more difficult to separate *underlying trends or patterns* from random and uninteresting features. Financial data are also almost always not normally distributed in spite of the fact that most techniques in econometrics assume that they are. High frequency data often contain additional ‘patterns’ which are the result of the way that the market works, or the way that prices are recorded. These features need to be considered in the model-building process, even if they are not directly of interest to the researcher.

### **1.3 Types of data**

There are broadly three types of data that can be employed in quantitative analysis of financial problems: time series data, cross-sectional data, and panel data.

#### *1.3.1 Time series data*

Time series data, as the name suggests, are data that have been collected over a period of time on one or more variables. Time series data have

**Box 1.2** Time series data

<i>Series</i>	<i>Frequency</i>
Industrial production	Monthly, or quarterly
Government budget deficit	Annually
Money supply	Weekly
The value of a stock	As transactions occur

associated with them a particular frequency of observation or collection of data points. The frequency is simply a measure of the *interval over*, or the *regularity with which*, the data are collected or recorded. Box 1.2 shows some examples of time series data.

A word on ‘As transactions occur’ is necessary. Much financial data does not start its life as being *regularly spaced*. For example, the price of common stock for a given company might be recorded to have changed whenever there is a new trade or quotation placed by the financial information recorder. Such recordings are very unlikely to be evenly distributed over time – for example, there may be no activity between, say, 5p.m. when the market closes and 8.30a.m. the next day when it reopens; there is also typically less activity around the opening and closing of the market, and around lunch time. Although there are a number of ways to deal with this issue, a common and simple approach is simply to select an appropriate frequency, and use as the observation for that time period the last prevailing price during the interval.

It is also generally a requirement that all data used in a model be of the *same frequency of observation*. So, for example, regressions that seek to estimate an arbitrage pricing model using monthly observations on macroeconomic factors must also use monthly observations on stock returns, even if daily or weekly observations on the latter are available.

The data may be *quantitative* (e.g. exchange rates, prices, number of shares outstanding), or *qualitative* (e.g. the day of the week, a survey of the financial products purchased by private individuals over a period of time, a credit rating, etc.).

**Problems that could be tackled using time series data:**

- How the value of a country’s stock index has varied with that country’s macroeconomic fundamentals
- How the value of a company’s stock price has varied when it announced the value of its dividend payment
- The effect on a country’s exchange rate of an increase in its trade deficit.

In all of the above cases, it is clearly the time dimension which is the most important, and the analysis will be conducted using the values of the variables over time.

### 1.3.2 Cross-sectional data

Cross-sectional data are data on one or more variables collected at a single point in time. For example, the data might be on:

- A poll of usage of Internet stockbroking services
- A cross-section of stock returns on the New York Stock Exchange (NYSE)
- A sample of bond credit ratings for UK banks.

#### **Problems that could be tackled using cross-sectional data:**

- The relationship between company size and the return to investing in its shares
- The relationship between a country's GDP level and the probability that the government will default on its sovereign debt.

### 1.3.3 Panel data

Panel data have the dimensions of both time series and cross-sections, e.g. the daily prices of a number of blue chip stocks over two years. The estimation of panel regressions is an interesting and developing area, and will be examined in detail in chapter 10.

Fortunately, virtually all of the standard techniques and analysis in econometrics are equally valid for time series and cross-sectional data. For time series data, it is usual to denote the individual observation numbers using the index  $t$ , and the total number of observations available for analysis by  $T$ . For cross-sectional data, the individual observation numbers are indicated using the index  $i$ , and the total number of observations available for analysis by  $N$ . Note that there is, in contrast to the time series case, no natural ordering of the observations in a cross-sectional sample. For example, the observations  $i$  might be on the price of bonds of different firms at a particular point in time, ordered alphabetically by company name. So, in the case of cross-sectional data, there is unlikely to be any useful information contained in the fact that Northern Rock follows National Westminster in a sample of UK bank credit ratings, since it is purely by chance that their names both begin with the letter 'N'. On the other hand, in a time series context, the ordering of the data is relevant since the data are usually ordered chronologically.

In this book, the total number of observations in the sample will be given by  $T$  even in the context of regression equations that could apply either to cross-sectional or to time series data.

#### 1.3.4 *Continuous and discrete data*

As well as classifying data as being of the time series or cross-sectional type, we could also distinguish it as being either continuous or discrete, exactly as their labels would suggest. *Continuous* data can take on any value and are not confined to take specific numbers; their values are limited only by precision. For example, the rental yield on a property could be 6.2%, 6.24% or 6.238%, and so on. On the other hand, *discrete* data can only take on certain values, which are usually integers<sup>1</sup> (whole numbers), and are often defined to be count numbers. For instance, the number of people in a particular underground carriage or the number of shares traded during a day. In these cases, having 86.3 passengers in the carriage or 5857 $\frac{1}{2}$  shares traded would not make sense.

#### 1.3.5 *Cardinal, ordinal and nominal numbers*

Another way in which we could classify numbers is according to whether they are cardinal, ordinal, or nominal. *Cardinal* numbers are those where the actual numerical values that a particular variable takes have meaning, and where there is an equal distance between the numerical values. On the other hand, *ordinal* numbers can only be interpreted as providing a position or an ordering. Thus, for cardinal numbers, a figure of 12 implies a measure that is 'twice as good' as a figure of 6. Examples of cardinal numbers would be the price of a share or of a building, and the number of houses in a street. On the other hand, for an ordinal scale, a figure of 12 may be viewed as 'better' than a figure of 6, but could not be considered twice as good. Examples of ordinal numbers would be the position of a runner in a race (e.g. second place is better than fourth place, but it would make little sense to say it is 'twice as good') or the level reached in a computer game.

The final type of data that could be encountered would be where there is no natural ordering of the values at all, so a figure of 12 is simply different to that of a figure of 6, but could not be considered to be better or worse in any sense. Such data often arise when numerical values are arbitrarily assigned, such as telephone numbers or when codings are assigned to

<sup>1</sup> Discretely measured data do not necessarily have to be integers. For example, until recently when they became 'decimalised', many financial asset prices were quoted to the nearest  $\frac{1}{16}$  or  $\frac{1}{32}$  of a dollar.

qualitative data (e.g. when describing the exchange that a US stock is traded on, '1' might be used to denote the NYSE, '2' to denote the NASDAQ and '3' to denote the AMEX). Sometimes, such variables are called *nominal* variables. Cardinal, ordinal and nominal variables may require different modelling approaches or at least different treatments, as should become evident in the subsequent chapters.

## 1.4 Returns in financial modelling

In many of the problems of interest in finance, the starting point is a time series of prices – for example, the prices of shares in Ford, taken at 4p.m. each day for 200 days. For a number of statistical reasons, it is preferable not to work directly with the price series, so that raw price series are usually converted into series of returns. Additionally, returns have the added benefit that they are unit-free. So, for example, if an annualised return were 10%, then investors know that they would have got back £110 for a £100 investment, or £1,100 for a £1,000 investment, and so on.

There are two methods used to calculate returns from a series of prices, and these involve the formation of simple returns, and continuously compounded returns, which are achieved as follows:

*Simple returns*

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}} \times 100\% \quad (1.1)$$

*Continuously compounded returns*

$$r_t = 100\% \times \ln \left( \frac{p_t}{p_{t-1}} \right) \quad (1.2)$$

where:  $R_t$  denotes the simple return at time  $t$ ,  $r_t$  denotes the continuously compounded return at time  $t$ ,  $p_t$  denotes the asset price at time  $t$ , and  $\ln$  denotes the natural logarithm.

If the asset under consideration is a stock or portfolio of stocks, the total return to holding it is the sum of the capital gain and any dividends paid during the holding period. However, researchers often ignore any dividend payments. This is unfortunate, and will lead to an underestimation of the total returns that accrue to investors. This is likely to be negligible for very short holding periods, but will have a severe impact on cumulative returns over investment horizons of several years. Ignoring dividends will also have a distortionary effect on the cross-section of stock returns. For example, ignoring dividends will imply that 'growth' stocks, with large capital gains will be inappropriately favoured over income stocks (e.g. utilities and mature industries) that pay high dividends.

**Box 1.3** Log returns

- (1) Log-returns have the nice property that they can be interpreted as *continuously compounded returns* – so that the frequency of compounding of the return does not matter and thus returns across assets can more easily be compared.
- (2) Continuously compounded returns are *time-additive*. For example, suppose that a weekly returns series is required and daily log returns have been calculated for five days, numbered 1 to 5, representing the returns on Monday through Friday. It is valid to simply add up the five daily returns to obtain the return for the whole week:

Monday return	$r_1 = \ln(p_1/p_0) = \ln p_1 - \ln p_0$
Tuesday return	$r_2 = \ln(p_2/p_1) = \ln p_2 - \ln p_1$
Wednesday return	$r_3 = \ln(p_3/p_2) = \ln p_3 - \ln p_2$
Thursday return	$r_4 = \ln(p_4/p_3) = \ln p_4 - \ln p_3$
Friday return	$r_5 = \ln(p_5/p_4) = \ln p_5 - \ln p_4$
Return over the week	$\ln p_5 - \ln p_0 = \ln(p_5/p_0)$

Alternatively, it is possible to adjust a stock price time series so that the dividends are added back to generate a *total return index*. If  $p_t$  were a total return index, returns generated using either of the two formulae presented above thus provide a measure of the total return that would accrue to a holder of the asset during time  $t$ .

The academic finance literature generally employs the log-return formulation (also known as log-price relatives since they are the log of the ratio of this period's price to the previous period's price). Box 1.3 shows two key reasons for this.

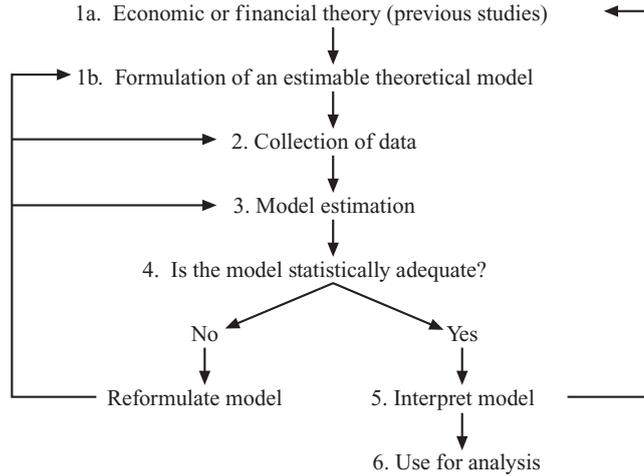
There is, however, also a disadvantage of using the log-returns. The simple return on a portfolio of assets is a weighted average of the simple returns on the individual assets:

$$R_{pt} = \sum_{i=1}^N w_i R_{it} \quad (1.3)$$

But this does not work for the continuously compounded returns, so that they are not additive across a portfolio. The fundamental reason why this is the case is that the log of a sum is not the same as the sum of a log, since the operation of taking a log constitutes a *non-linear transformation*. Calculating portfolio returns in this context must be conducted by first estimating the value of the portfolio at each time period and then determining the returns from the aggregate portfolio values. Or alternatively, if we assume that the asset is purchased at time  $t - K$  for price  $P_{t-K}$  and then sold  $K$  periods later at price  $P_t$ , then if we calculate simple returns for each period,  $R_t, R_{t+1}, \dots, R_K$ , the aggregate return over all  $K$

**Figure 1.1**

Steps involved in forming an econometric model



periods is

$$\begin{aligned}
 R_{Kt} &= \frac{P_t - P_{t-K}}{P_{t-K}} = \frac{P_t}{P_{t-K}} - 1 = \left[ \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \dots \times \frac{P_{t-K+1}}{P_{t-K}} \right] - 1 \\
 &= [(1 + R_t)(1 + R_{t-1}) \dots (1 + R_{t-K+1})] - 1
 \end{aligned}
 \tag{1.4}$$

In the limit, as the frequency of the sampling of the data is increased so that they are measured over a smaller and smaller time interval, the simple and continuously compounded returns will be identical.

## 1.5 Steps involved in formulating an econometric model

Although there are of course many different ways to go about the process of model building, a logical and valid approach would be to follow the steps described in figure 1.1.

The steps involved in the model construction process are now listed and described. Further details on each stage are given in subsequent chapters of this book.

- *Step 1a and 1b: general statement of the problem* This will usually involve the formulation of a theoretical model, or intuition from financial theory that two or more variables should be related to one another in a certain way. The model is unlikely to be able to completely capture every relevant real-world phenomenon, but it should present a sufficiently good approximation that it is useful for the purpose at hand.

- *Step 2: collection of data relevant to the model* The data required may be available electronically through a financial information provider, such as Reuters or from published government figures. Alternatively, the required data may be available only via a survey after distributing a set of questionnaires i.e. *primary data*.
- *Step 3: choice of estimation method relevant to the model proposed in step 1* For example, is a single equation or multiple equation technique to be used?
- *Step 4: statistical evaluation of the model* What assumptions were required to estimate the parameters of the model optimally? Were these assumptions satisfied by the data or the model? Also, does the model adequately describe the data? If the answer is ‘yes’, proceed to step 5; if not, go back to steps 1–3 and either reformulate the model, collect more data, or select a different estimation technique that has less stringent requirements.
- *Step 5: evaluation of the model from a theoretical perspective* Are the parameter estimates of the sizes and signs that the theory or intuition from step 1 suggested? If the answer is ‘yes’, proceed to step 6; if not, again return to stages 1–3.
- *Step 6: use of model* When a researcher is finally satisfied with the model, it can then be used for testing the theory specified in step 1, or for formulating forecasts or suggested courses of action. This suggested course of action might be for an individual (e.g. ‘if inflation and GDP rise, buy stocks in sector X’), or as an input to government policy (e.g. ‘when equity markets fall, program trading causes excessive volatility and so should be banned’).

It is important to note that the process of building a robust empirical model is an iterative one, and it is certainly not an exact science. Often, the final preferred model could be very different from the one originally proposed, and need not be unique in the sense that another researcher with the same data and the same initial theory could arrive at a different final specification.

## 1.6 Points to consider when reading articles in empirical finance

As stated above, one of the defining features of this book relative to others in the area is in its use of published academic research as examples of the use of the various techniques. The papers examined have been chosen for a number of reasons. Above all, they represent (in this author’s opinion) a clear and specific application in finance of the techniques covered in this

**Box 1.4** Points to consider when reading a published paper

- (1) Does the paper involve the development of a theoretical model or is it merely a technique looking for an application so that the motivation for the whole exercise is poor?
- (2) Are the data of 'good quality'? Are they from a reliable source? Is the size of the sample sufficiently large for the model estimation task at hand?
- (3) Have the techniques been validly applied? Have tests been conducted for possible violations of any assumptions made in the estimation of the model?
- (4) Have the results been interpreted sensibly? Is the strength of the results exaggerated? Do the results actually obtained relate to the questions posed by the author(s)? Can the results be replicated by other researchers?
- (5) Are the conclusions drawn appropriate given the results, or has the importance of the results of the paper been overstated?

book. They were also required to be published in a peer-reviewed journal, and hence to be widely available.

When I was a student, I used to think that research was a very pure science. Now, having had first-hand experience of research that academics and practitioners do, I know that this is not the case. Researchers often cut corners. They have a tendency to exaggerate the strength of their results, and the importance of their conclusions. They also have a tendency not to bother with tests of the adequacy of their models, and to gloss over or omit altogether any results that do not conform to the point that they wish to make. Therefore, when examining papers from the academic finance literature, it is important to cast a very critical eye over the research – rather like a referee who has been asked to comment on the suitability of a study for a scholarly journal. The questions that are always worth asking oneself when reading a paper are outlined in box 1.4.

Bear these questions in mind when reading my summaries of the articles used as examples in this book and, if at all possible, seek out and read the entire articles for yourself.

## 1.7 Econometric packages for modelling financial data

As the name suggests, this section contains descriptions of various computer packages that may be employed to estimate econometric models. The number of available packages is large, and over time, all packages have improved in breadth of available techniques, and have also converged in terms of what is available in each package. Some readers may already be familiar with the use of one or more packages, and if this is the case, this section may be skipped. For those who do not know how to use any

**Table 1.1** Econometric software packages for modelling financial data

	Package software supplier*
EViews	QMS Software
GAUSS	Aptech Systems
LIMDEP	Econometric Software
MATLAB	The MathWorks
RATS	Estima
SAS	SAS Institute
SHAZAM	Northwest Econometrics
SPLUS	Insightful Corporation
SPSS	SPSS
TSP	TSP International

\* Full contact details for all software suppliers can be found in the appendix at the end of this chapter.

econometrics software, or have not yet found a package which suits their requirements, then read on.

### 1.7.1 *What packages are available?*

Although this list is by no means exhaustive, a set of widely used packages is given in table 1.1. The programs can usefully be categorised according to whether they are fully interactive, (menu-driven), command-driven (so that the user has to write mini-programs), or somewhere in between. Menu-driven packages, which are usually based on a standard Microsoft Windows graphical user interface, are almost certainly the easiest for novices to get started with, for they require little knowledge of the structure of the package, and the menus can usually be negotiated simply. EViews is a package that falls into this category.

On the other hand, some such packages are often the least flexible, since the menus of available options are fixed by the developers, and hence if one wishes to build something slightly more complex or just different, then one is forced to consider alternatives. EViews, however, has a command-based programming language as well as a click-and-point interface so that it offers flexibility as well as user-friendliness.

### 1.7.2 *Choosing a package*

Choosing an econometric software package is an increasingly difficult task as the packages become more powerful but at the same time more homogeneous. For example, LIMDEP, a package originally developed for the analysis of a certain class of cross-sectional data, has many useful

features for modelling financial time series. Also, many packages developed for time series analysis, such as TSP (“Time Series Processor”), can also now be used for cross-sectional or panel data. Of course, this choice may be made for you if your institution offers or supports only one or two of the above possibilities. Otherwise, sensible questions to ask yourself are:

- Is the package suitable for your *intended applications* – for example, does the software have the capability for the models that you want to estimate? Can it handle sufficiently large databases?
- Is the package *user-friendly*?
- Is it *fast*?
- How much does it *cost*?
- Is it *accurate*?
- Is the package *discussed* or *supported* in a standard textbook, as EViews is in this book?
- Does the package have *readable and comprehensive manuals*? Is help available online?
- Does the package come with *free technical support* so that you can e-mail the developers with queries?

A great deal of useful information can be obtained most easily from the web pages of the software developers. Additionally, many journals (including the *Journal of Applied Econometrics*, the *Economic Journal*, the *International Journal of Forecasting* and the *American Statistician*) publish software reviews that seek to evaluate and compare the packages’ usefulness for a given purpose. Three reviews that this author has been involved with, that are relevant for chapter 8 of this text in particular, are Brooks (1997) and Brooks, Burke and Persaud (2001, 2003).

The EViews package will be employed in this text because it is simple to use, menu-driven, and will be sufficient to estimate most of the models required for this book. The following section gives an introduction to this software and outlines the key features and how basic tasks are executed.<sup>2</sup>

### 1.7.3 Accomplishing simple tasks using EViews

EViews is a simple to use, interactive econometrics software package, providing the tools most frequently used in practical econometrics. EViews is built around the concept of objects with each object having its own window, its own menu, its own procedure and its own view of its data.

<sup>2</sup> The first edition of this text also incorporated a detailed discussion of the WinRATS package, but in the interests of keeping the book at a manageable length with two new chapters included, the support for WinRATS users will now be given in a separate handbook that accompanies the main text, ISBN: 9780521896955.

Using menus, it is easy to change between displays of a spreadsheet, line and bar graphs, regression results, etc. One of the most important features of EViews that makes it useful for model-building is the wealth of diagnostic (misspecification) tests, that are automatically computed, making it possible to test whether the model is econometrically valid or not. You work your way through EViews using a combination of windows, buttons, menus and sub-menus. A good way of familiarising yourself with EViews is to learn about its main menus and their relationships through the examples given in this and subsequent chapters.

This section assumes that readers have obtained a licensed copy of EViews, and have successfully loaded it onto an available computer. There now follows a description of the EViews package, together with instructions to achieve standard tasks and sample output. Any instructions that must be entered or icons to be clicked are illustrated throughout this book by **bold-faced type**. The objective of the treatment in this and subsequent chapters is not to demonstrate the full functionality of the package, but rather to get readers started quickly and to explain how the techniques are implemented. For further details, readers should consult the software manuals in the first instance, which are now available electronically with the software as well as in hard copy.<sup>3</sup> Note that EViews is not case-sensitive, so that it does not matter whether commands are entered as lower-case or CAPITAL letters.

#### Opening the software

To load EViews from Windows, choose **Start, All Programs, EViews6** and finally, **EViews6** again.

#### Reading in data

EViews provides support to read from or write to various file types, including 'ASCII' (text) files, Microsoft Excel '.XLS' files (reading from any named sheet in the Excel workbook), Lotus '.WKS1' and '.WKS3' files. It is usually easiest to work directly with Excel files, and this will be the case throughout this book.

#### Creating a workfile and importing data

The first step when the EViews software is opened is to create a *workfile* that will hold the data. To do this, select **New** from the File menu. Then

<sup>3</sup> A student edition of EViews 4.1 is available at a much lower cost than the full version, but with reduced functionality and restrictions on the number of observations and objects that can be included in each workfile.

choose **Workfile**. The ‘Workfile Create’ window in screenshot 1.1 will be displayed.

**Screenshot 1.1**  
Creating a workfile

We are going to use as an example a time series of UK average house price data obtained from Nationwide,<sup>4</sup> which comprises 197 monthly observations from January 1991 to May 2007. The frequency of the data (**Monthly**) should be set and the start (**1991:01**) and end (**2007:05**) dates should be inputted. Click **OK**. An untitled workfile will be created.

Under ‘Workfile structure type’, keep the default option, **Dated – regular frequency**. Then, under ‘Date specification’, choose **Monthly**. Note the format of date entry for monthly and quarterly data: **YYYY:M** and **YYYY:Q**, respectively. For daily data, a US date format must usually be used depending on how EViews has been set up: **MM/DD/YYYY** (e.g. 03/01/1999 would be 1st March 1999, not 3rd January). Caution therefore needs to be exercised here to ensure that the date format used is the correct one. Type the start and end dates for the sample into the boxes: **1991:01** and **2007:05** respectively. Then click **OK**. The workfile will now have been created. Note that two pairs of dates are displayed, ‘Range’ and ‘Sample’: the first one is the range of dates contained in the workfile and the second one (which is the same as above in this case) is for the current workfile sample. Two

<sup>4</sup> Full descriptions of the sources of data used will be given in appendix 3 and on the web site accompanying this book.

objects are also displayed: C (which is a vector that will eventually contain the parameters of any estimated models) and RESID (a residuals series, which will currently be empty). See chapter 2 for a discussion of these concepts. All EViews workfiles will contain these two objects, which are created automatically.

Now that the workfile has been set up, we can import the data from the Excel file UKHP.XLS. So from the File menu, select **Import** and **Read Text-Lotus-Excel**. You will then be prompted to select the directory and file name. Once you have found the directory where the file is stored, enter UKHP.XLS in the 'file name' box and select the file type 'Excel (\*.xls)'. The window in screenshot 1.2 ('Excel Spreadsheet Import') will be displayed.

### Screenshot 1.2

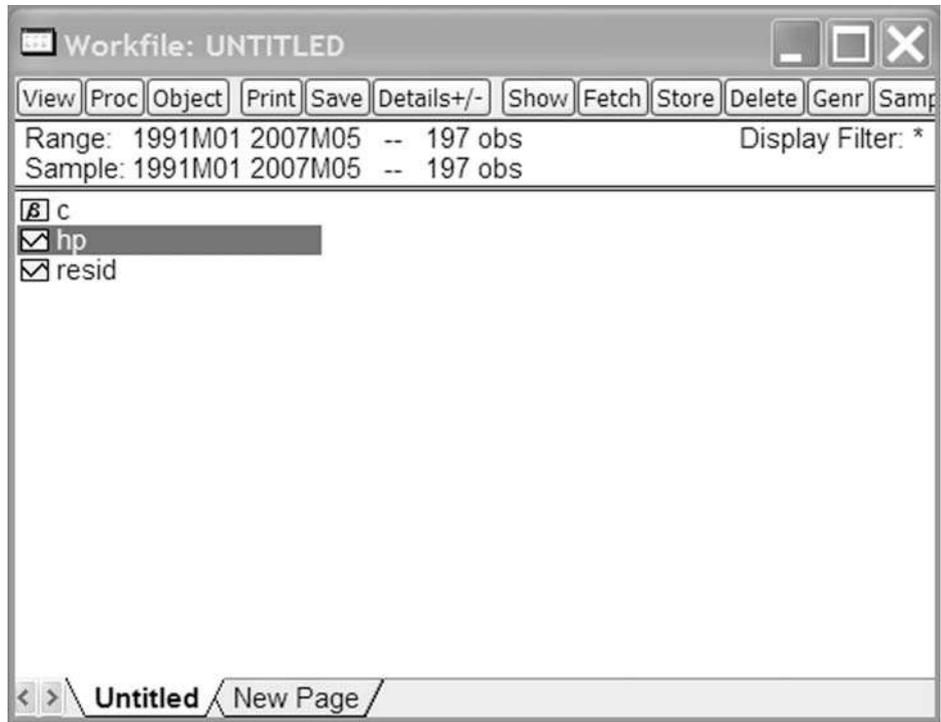
Importing Excel data into the workfile

You have to choose the order of your data: *by observations* (series in columns as they are in this and most other cases) or *by series* (series in rows). Also you could provide the names for your series in the relevant box. If the names of the series are already in the imported Excel data file, you can simply enter the number of series (which you are importing) in the 'Names for series or Number if named in file' field in the dialog box. In this case, enter **HP**, say, for house prices. The 'Upper-left data cell' refers to the first cell in the spreadsheet that actually contains numbers. In this case, it can be left at B2 as the first column in the spreadsheet contains

only dates and we do not need to import those since EViews will date the observations itself. You should also choose the sample of the data that you wish to import. This box can almost always be left at EViews' suggestion which defaults to the current workfile sample. Click OK and the series will be imported. The series will appear as a new icon in the workfile window, as in screenshot 1.3.

### Screenshot 1.3

The workfile containing loaded data



### Verifying the data

Double click on the new hp icon that has appeared, and this will open up a spreadsheet window within EViews containing the monthly house price values. Make sure that the data file has been correctly imported by checking a few observations at random.

The next step is to save the workfile: click on the **Save As** button from the **File** menu and select **Save Active Workfile** and click **OK**. A save dialog box will open, prompting you for a workfile name and location. You should enter XX (where XX is your chosen name for the file), then click **OK**. EViews will save the workfile in the specified directory with the name XX.WF1. The saved workfile can be opened later by selecting **File/Open/EViews Workfile . . .** from the menu bar.

### Transformations

Variables of interest can be created in EViews by selecting the *Genr* button from the workfile toolbar and typing in the relevant formulae. Suppose, for example, we have a time series called *Z*. The latter can be modified in the following ways so as to create Variables *A*, *B*, *C*, etc.

$A = Z/2$	Dividing
$B = Z*2$	Multiplication
$C = Z^2$	Squaring
$D = \text{LOG}(Z)$	Taking the logarithms
$E = \text{EXP}(Z)$	Taking the exponential
$F = Z(-1)$	Lagging the data
$G = \text{LOG}(Z/Z(-1))$	Creating the log-returns

Other functions that can be used in the formulae include: *abs*, *sin*, *cos*, etc. Notice that no special instruction is necessary; simply type ‘new variable = function of old variable(s)’. The variables will be displayed in the same workfile window as the original (imported) series.

In this case, it is of interest to calculate simple percentage changes in the series. Click **Genr** and type  $\text{DHP} = 100 * (\text{HP} - \text{HP}(-1)) / \text{HP}(-1)$ . It is important to note that this new series, *DHP*, will be a series of monthly changes and will not be annualised.

### Computing summary statistics

Descriptive summary statistics of a series can be obtained by selecting **Quick/Series Statistics/Histogram and Stats** and typing in the name of the variable (*DHP*). The view in screenshot 1.4 will be displayed in the window.

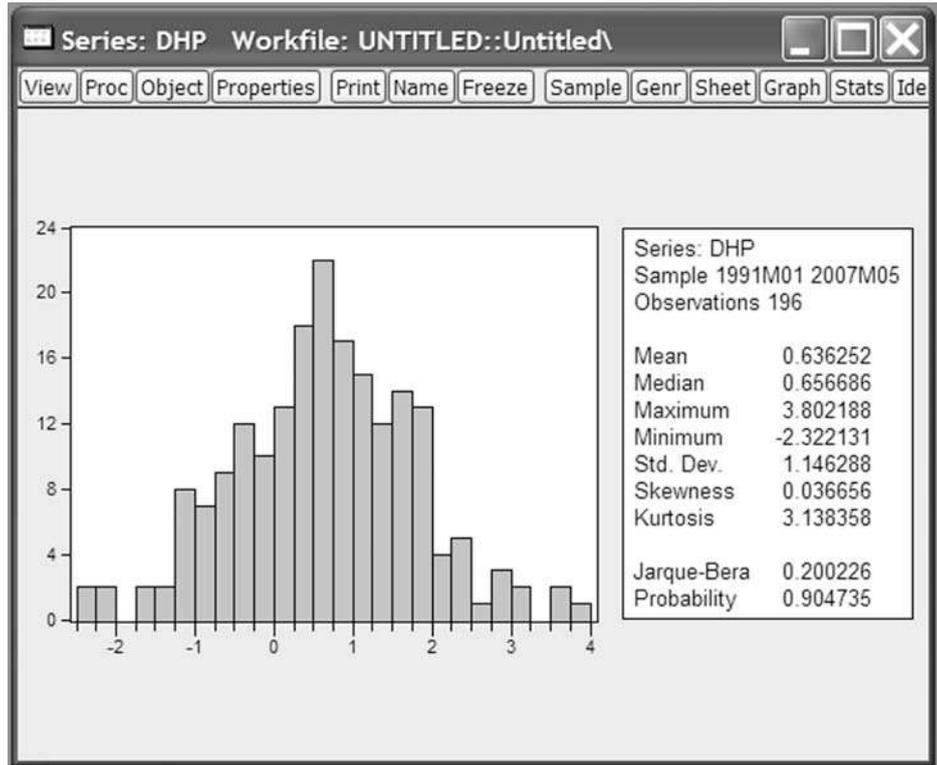
As can be seen, the histogram suggests that the series has a longer upper tail than lower tail (note the x-axis scale) and is centred slightly above zero. Summary statistics including the mean, maximum and minimum, standard deviation, higher moments and a test for whether the series is normally distributed are all presented. Interpreting these will be discussed in subsequent chapters. Other useful statistics and transformations can be obtained by selecting the command *Quick/Series Statistics*, but these are covered later in this book.

### Plots

EViews supports a wide range of graph types including line graphs, bar graphs, pie charts, mixed line–bar graphs, high–low graphs and scatter-plots. A variety of options permits the user to select the line types, colour,

**Screenshot 1.4**

Summary statistics  
for a series



border characteristics, headings, shading and scaling, including logarithmic scale and dual scale graphs. Legends are automatically created (although they can be removed if desired), and customised graphs can be incorporated into other Windows applications using copy-and-paste, or by exporting as Windows metafiles.

From the main menu, select **Quick/Graph** and type in the name of the series that you want to plot (**HP** to plot the level of house prices) and click **OK**. You will be prompted with the Graph window where you choose the type of graph that you want (line, bar, scatter or pie charts). There is a *Show Option* button, which you click to make adjustments to the graphs. Choosing a line graph would produce screenshot 1.5.

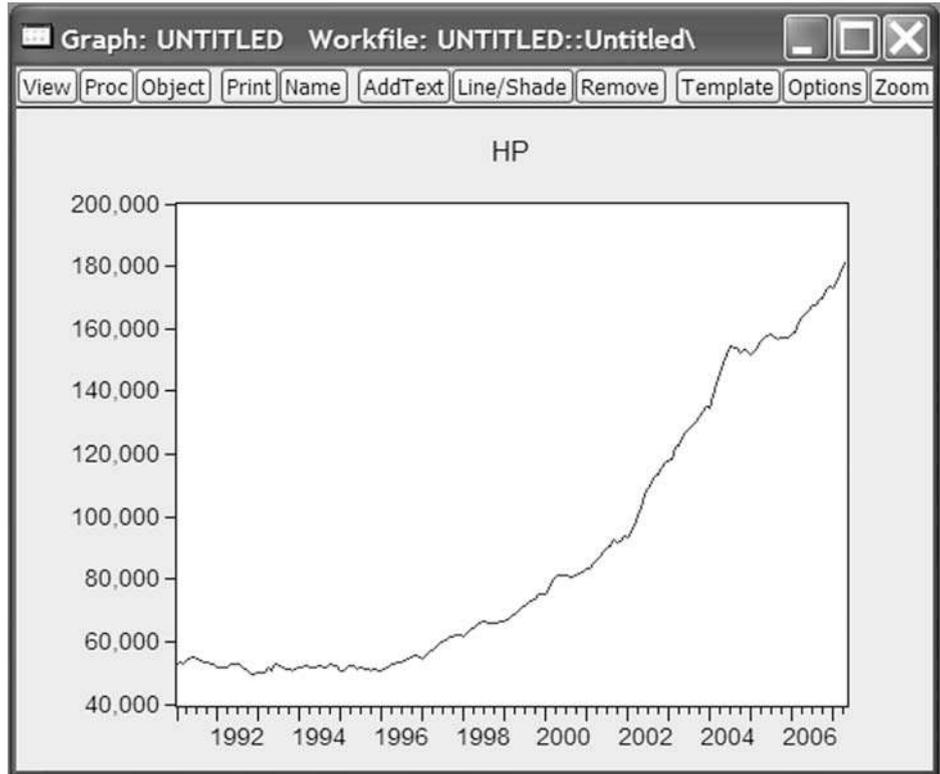
Scatter plots can similarly be produced by selecting 'Scatter' in the 'Graph Type' box after opening a new graph object.

### Printing results

Results can be printed at any point by selecting the *Print* button on the object window toolbar. The whole current window contents will be printed. Choosing *View/Print Selected* from the workfile window prints the default

**Screenshot 1.5**

A line graph



view for all of the selected objects. Graphs can be copied into the clipboard if desired by right clicking on the graph and choosing *Copy*.

#### Saving data results and workfile

Data generated in EViews can be exported to other Windows applications, e.g. Microsoft Excel. From the object toolbar, select *Procs/Export/Write Text-Lotus-Excel*. You will then be asked to provide a name for the exported file and to select the appropriate directory. The next window will ask you to select all the series that you want to export, together with the sample period.

Assuming that the workfile has been saved after the importation of the data set (as mentioned above), additional work can be saved by just selecting *Save* from the *File* menu. It will ask you if you want to overwrite the existing file, in which case you click on the *Yes* button. You will also be prompted to select whether the data in the file should be saved in 'single precision' or 'double precision'. The latter is preferable for obvious reasons unless the file is likely to be very large because of the quantity of variables and observations it contains (single precision will require less space). The workfile will be saved including all objects in it – data, graphs,

equations, etc. *so long as they have been given a title*. Any untitled objects will be lost upon exiting the program.

Econometric tools available in EViews

Box 1.5 describes the features available in EViews, following the format of the user guides for version 6, with material discussed in this book indicated by *italics*.

### Box 1.5 Features of EViews

The EViews user guide is now split into two volumes. Volume I contains parts I to III as described below, while Volume II contains Parts IV to VIII.

#### PART I (EViews FUNDAMENTALS)

- Chapters 1–4 contain introductory material describing the basics of Windows and EViews, how workfiles are constructed and how to deal with objects.
- Chapters 5 and 6 document the basics of working with data. *Importing data* into EViews, *using EViews to manipulate and manage data*, and exporting from EViews into spreadsheets, text files and other Windows applications are discussed.
- Chapters 7–10 describe the EViews database and other advanced data and workfile handling features.

#### PART II (BASIC DATA ANALYSIS)

- Chapter 11 describes the series object. Series are the basic unit of data in EViews and are the basis for all *univariate analysis*. This chapter documents the basic *graphing* and *data analysis features associated with series*.
- Chapter 12 documents the group object. Groups are collections of series that form the basis for a variety of *multivariate graphing* and *data analyses*.
- Chapter 13 provides detailed documentation for explanatory data analysis using distribution graphs, *density plots* and *scatter plot graphs*.
- Chapters 14 and 15 describe the creation and customisation of more advanced tables and graphs.

#### PART III (COMMANDS AND PROGRAMMING)

- Chapters 16–23 describe in detail how to write programs using the EViews programming language.

#### PART IV (BASIC SINGLE EQUATION ANALYSIS)

- Chapter 24 outlines the basics of *ordinary least squares estimation (OLS)* in EViews.
- Chapter 25 discusses the *weighted least squares*, *two-stage least squares* and non-linear least squares estimation techniques.
- Chapter 26 describes *single equation regression techniques* for the analysis of time series data: *testing for serial correlation*, *estimation of ARMA models*, using polynomial distributed lags, and *unit root tests* for non-stationary time series.

- Chapter 27 describes the fundamentals of using EViews to *forecast from estimated equations*.
- Chapter 28 describes the *specification testing* procedures available in EViews.

#### PART V (ADVANCED SINGLE EQUATION ANALYSIS)

- Chapter 29 discusses *ARCH and GARCH estimation* and outlines the EViews tools for *modelling the conditional variance* of a variable.
- Chapter 30 documents EViews functions for estimating *qualitative and limited dependent variable models*. EViews provides estimation routines for *binary or ordered (e.g. probit and logit), censored or truncated (tobit, etc.)* and integer valued (count) data.
- Chapter 31 discusses the fashionable topic of the estimation of quantile regressions.
- Chapter 32 shows how to deal with the log-likelihood object, and how to solve problems with non-linear estimation.

#### PART VI (MULTIPLE EQUATION ANALYSIS)

- Chapters 33–36 describe *estimation techniques for systems of equations* including VAR and VEC models, and state space models.

#### PART VII (PANEL AND POOLED DATA)

- Chapter 37 outlines tools for working with *pooled time series, cross-section data* and estimating standard equation specifications that account for the pooled structure of the data.
- Chapter 38 describes how to structure a panel of data and how to analyse it, while chapter 39 extends the analysis to look at *panel regression model estimation*.

#### PART VIII (OTHER MULTIVARIATE ANALYSIS)

- Chapter 40, the final chapter of the manual, explains how to conduct factor analysis in EViews.

## 1.8 Outline of the remainder of this book

### Chapter 2

This introduces the classical linear regression model (CLRM). The ordinary least squares (OLS) estimator is derived and its interpretation discussed. The conditions for OLS optimality are stated and explained. A hypothesis testing framework is developed and examined in the context of the linear model. Examples employed include Jensen's classic study of mutual fund performance measurement and tests of the 'overreaction hypothesis' in the context of the UK stock market.

### *Chapter 3*

This continues and develops the material of chapter 2 by generalising the bivariate model to multiple regression – i.e. models with many variables. The framework for testing multiple hypotheses is outlined, and measures of how well the model fits the data are described. Case studies include modelling rental values and an application of principal components analysis to interest rate modelling.

### *Chapter 4*

Chapter 4 examines the important but often neglected topic of diagnostic testing. The consequences of violations of the CLRM assumptions are described, along with plausible remedial steps. Model-building philosophies are discussed, with particular reference to the general-to-specific approach. Applications covered in this chapter include the determination of sovereign credit ratings.

### *Chapter 5*

This presents an introduction to time series models, including their motivation and a description of the characteristics of financial data that they can and cannot capture. The chapter commences with a presentation of the features of some standard models of stochastic (white noise, moving average, autoregressive and mixed ARMA) processes. The chapter continues by showing how the appropriate model can be chosen for a set of actual data, how the model is estimated and how model adequacy checks are performed. The generation of forecasts from such models is discussed, as are the criteria by which these forecasts can be evaluated. Examples include model-building for UK house prices, and tests of the exchange rate covered and uncovered interest parity hypotheses.

### *Chapter 6*

This extends the analysis from univariate to multivariate models. Multivariate models are motivated by way of explanation of the possible existence of bi-directional causality in financial relationships, and the simultaneous equations bias that results if this is ignored. Estimation techniques for simultaneous equations models are outlined. Vector autoregressive (VAR) models, which have become extremely popular in the empirical finance literature, are also covered. The interpretation of VARs is explained by way of joint tests of restrictions, causality tests, impulse responses and variance decompositions. Relevant examples discussed in this chapter are the simultaneous relationship between bid-ask spreads

and trading volume in the context of options pricing, and the relationship between property returns and macroeconomic variables.

### *Chapter 7*

The first section of the chapter discusses unit root processes and presents tests for non-stationarity in time series. The concept of and tests for cointegration, and the formulation of error correction models, are then discussed in the context of both the single equation framework of Engle-Granger, and the multivariate framework of Johansen. Applications studied in chapter 7 include spot and futures markets, tests for cointegration between international bond markets and tests of the purchasing power parity hypothesis and of the expectations hypothesis of the term structure of interest rates.

### *Chapter 8*

This covers the important topic of volatility and correlation modelling and forecasting. This chapter starts by discussing in general terms the issue of non-linearity in financial time series. The class of ARCH (AutoRegressive Conditionally Heteroscedastic) models and the motivation for this formulation are then discussed. Other models are also presented, including extensions of the basic model such as GARCH, GARCH-M, EGARCH and GJR formulations. Examples of the huge number of applications are discussed, with particular reference to stock returns. Multivariate GARCH models are described, and applications to the estimation of conditional betas and time-varying hedge ratios, and to financial risk measurement, are given.

### *Chapter 9*

This discusses testing for and modelling regime shifts or switches of behaviour in financial series that can arise from changes in government policy, market trading conditions or microstructure, among other causes. This chapter introduces the Markov switching approach to dealing with regime shifts. Threshold autoregression is also discussed, along with issues relating to the estimation of such models. Examples include the modelling of exchange rates within a managed floating environment, modelling and forecasting the gilt-equity yield ratio, and models of movements of the difference between spot and futures prices.

### *Chapter 10*

This new chapter focuses on how to deal appropriately with longitudinal data – that is, data having both time series and cross-sectional dimensions. Fixed effect and random effect models are explained and illustrated by way

of examples on banking competition in the UK and on credit stability in Central and Eastern Europe. Entity fixed and time-fixed effects models are elucidated and distinguished.

### *Chapter 11*

The second new chapter describes various models that are appropriate for situations where the dependent variable is not continuous. Readers will learn how to construct, estimate and interpret such models, and to distinguish and select between alternative specifications. Examples used include a test of the pecking order hypothesis in corporate finance and the modelling of unsolicited credit ratings.

### *Chapter 12*

This presents an introduction to the use of simulations in econometrics and finance. Motivations are given for the use of repeated sampling, and a distinction is drawn between Monte Carlo simulation and bootstrapping. The reader is shown how to set up a simulation, and examples are given in options pricing and financial risk management to demonstrate the usefulness of these techniques.

### *Chapter 13*

This offers suggestions related to conducting a project or dissertation in empirical finance. It introduces the sources of financial and economic data available on the Internet and elsewhere, and recommends relevant online information and literature on research in financial markets and financial time series. The chapter also suggests ideas for what might constitute a good structure for a dissertation on this subject, how to generate ideas for a suitable topic, what format the report could take, and some common pitfalls.

### *Chapter 14*

This summarises the book and concludes. Several recent developments in the field, which are not covered elsewhere in the book, are also mentioned. Some tentative suggestions for possible growth areas in the modelling of financial time series are also given.

## **1.9 Further reading**

*EViews 6 User's Guides I and II* – Quantitative Micro Software (2007), QMS, Irvine, CA  
*EViews 6 Command Reference* – Quantitative Micro Software (2007), QMS, Irvine, CA  
Startz, R. *EViews Illustrated for Version 6* (2007) QMS, Irvine, CA

**Appendix: Econometric software package suppliers**

Package	Contact information
EViews	QMS Software, Suite 336, 4521 Campus Drive #336, Irvine, CA 92612-2621, USA Tel: (+1) 949 856 3368; Fax: (+1) 949 856 2044; Web: <a href="http://www.eviews.com">www.eviews.com</a>
GAUSS	Aptech Systems Inc, PO Box 250, Black Diamond, WA 98010, USA Tel: (+1) 425 432 7855; Fax: (+1) 425 432 7832; Web: <a href="http://www.aptech.com">www.aptech.com</a>
LIMDEP	Econometric Software, 15 Gloria Place, Plainview, NY 11803, USA Tel: (+1) 516 938 5254; Fax: (+1) 516 938 2441; Web: <a href="http://www.limdep.com">www.limdep.com</a>
MATLAB	The MathWorks Inc., 3 Applie Hill Drive, Natick, MA 01760-2098, USA Tel: (+1) 508 647 7000; Fax: (+1) 508 647 7001; Web: <a href="http://www.mathworks.com">www.mathworks.com</a>
RATS	Estima, 1560 Sherman Avenue, Evanston, IL 60201, USA Tel: (+1) 847 864 8772; Fax: (+1) 847 864 6221; Web: <a href="http://www.estima.com">www.estima.com</a>
SAS	SAS Institute, 100 Campus Drive, Cary NC 27513-2414, USA Tel: (+1) 919 677 8000; Fax: (+1) 919 677 4444; Web: <a href="http://www.sas.com">www.sas.com</a>
SHAZAM	Northwest Econometrics Ltd., 277 Arbutus Reach, Gibsons, B.C. V0N 1V8, Canada Tel: -; Fax: (+1) 707 317 5364; Web: <a href="http://shazam.econ.ubc.ca">shazam.econ.ubc.ca</a>
SPLUS	Insightful Corporation, 1700 Westlake Avenue North, Suite 500, Seattle, WA 98109-3044, USA Tel: (+1) 206 283 8802; Fax: (+1) 206 283 8691; Web: <a href="http://www.splus.com">www.splus.com</a>
SPSS	SPSS Inc, 233 S. Wacker Drive, 11th Floor, Chicago, IL 60606-6307, USA Tel: (+1) 800 543 2185; Fax: (+1) 800 841 0064; Web: <a href="http://www.spss.com">www.spss.com</a>
TSP	TSP International, PO Box 61015 Station A, Palo Alto, CA 94306, USA Tel: (+1) 650 326 1927; Fax: (+1) 650 328 4163; Web: <a href="http://www.tspintl.com">www.tspintl.com</a>

**Key concepts**

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

- financial econometrics
- time series
- panel data
- continuous data
- continuously compounded returns
- cross-sectional data
- pooled data
- discrete data