Introductory Econometrics for Finance

Chris Brooks
The ISMA Centre, University of Reading

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York, NY 10011-4211, USA
477 Williamstown Road, Port Melbourne, VIC 3207, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2002

Printed in the United Kingdom at the University Press, Cambridge

Typefaces Swift 10.5/14 pt. and Franklin Gothic *System* $\LaTeX\,2_\varepsilon$ [TB]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data
Brooks, Chris
Introductory econometrics for finance / Chris Brooks.
p. cm.
Includes bibliographical references and index.
ISBN 0 521 79018 2 (hardback) – ISBN 0 521 79367 X (paperback)
1. Finance – Econometric models. 2. Econometrics. I. Title.
HG173 .B76 2002
332′.01′5195 – dc21 2001037930

ISBN 0 521 79018 2 hardback
ISBN 0 521 79367 X paperback

Contents

Figures

Tables

Boxes

Screenshots

1 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.

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.

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.

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.

1.2 Is financial econometrics different from 'economic econometrics'? Some stylised characteristics of financial data

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 25 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 macroeconometricians! 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 has 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 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 is collected or recorded. Box 1.2 shows some examples of time series data.

Box 1.2 Time series data

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

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).

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 regression 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, but will not be considered further. Interested readers are referred to the excellent text by Baltagi (1995).

Fortunately, virtually all of the standard techniques and analysis in econometrics are equally valid for time series and cross-sectional data. This book will, however, concentrate mainly on time series data and applications since these are more prevalent in finance. 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.

1.4 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 *Continuously compounded returns*

$$R_t = \frac{p_t - p_{t-1}}{p_{t-1}} \times 100\% \quad (1.1) \qquad r_t = 100\% \times \ln\left(\frac{p_t}{p_{t-1}}\right) \quad (1.2)$$

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
ln denotes the natural logarithm

If the asset under consideration is a stock or portfolio of stocks, the total return to holding the stock 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 preferred over income stocks (e.g. utilities and mature industries) that pay high dividends.

Alternatively, it is often assumed that stock price time series have been adjusted and the dividends added back in to generate a total return index. Returns generated using either of the two formulae presented above thus prove 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:

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

But this does not work for the continuously compounded returns, so that they are not additive across a portfolio. The

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 5 days, numbered 1 to 5, representing the returns on Monday through Friday. It is valid to simply add up the 5 daily returns to obtain the return for the whole week:

$$\begin{array}{lll}
\textit{Monday return} & r_1 = \ln(p_1/p_0) = \ln p_1 - \ln p_0 & (1.4)\\
\textit{Tuesday return} & r_2 = \ln(p_2/p_1) = \ln p_2 - \ln p_1 & (1.5)\\
\textit{Wednesday return} & r_3 = \ln(p_3/p_2) = \ln p_3 - \ln p_2 & (1.6)\\
\textit{Thursday return} & r_4 = \ln(p_4/p_3) = \ln p_4 - \ln p_3 & (1.7)\\
\textit{Friday return} & r_5 = \ln(p_5/p_4) = \ln p_5 - \ln p_4 & (1.8)\\
\hline
\textit{Return over the week} & \ln p_5 - \ln p_0 = \ln(p_5/p_0) & (1.9)
\end{array}$$

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.

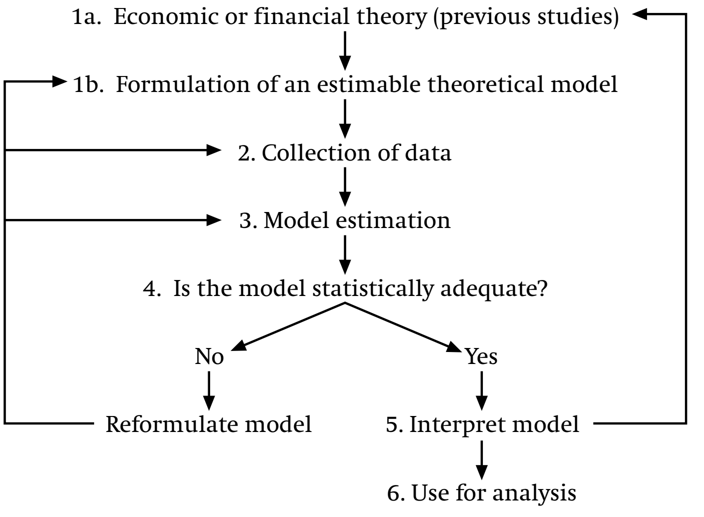
In the limit, as the frequency of the sampling of the data is increased, so that they are measured over a smaller and smaller time interval, the simple and continuously compounded returns will be identical.

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.

Figure 1.1 Steps involved in forming an econometric model



- *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, Bridge Telerate, or Primark Datastream, or from published government figures. Alternatively, the required data may be available only via a survey after distributing a set of questionnaires.
- *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 Some points to consider when reading articles in the empirical finance literature

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 in this book 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 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 paper – rather like a referee who has been asked to comment on the suitability of a paper for a scholarly journal.

The questions that are always worth asking oneself when reading a paper are outlined in box 1.4.

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. Is the data of 'good quality'? Is it 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?

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 Outline of the remainder of this book

Chapter 2

This gives contact details for a large number of econometrics packages which can be used for the modelling of financial time series, together with a description of two packages that will be examined in detail in this text (EViews and RATS). Brief introductions to the use of the packages for reading in data, plotting graphs, obtaining summary statistics, doing simple transformations, computing correlations, and so on, are also given.

Chapter 3

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. Single and multiple hypothesis testing frameworks are developed and examined in the context of the linear model. Examples employed include tests of the 'overreaction hypothesis' in the context of the UK stock market.

Chapter 4

This continues and develops the material of chapter 3 to consider goodness of fit statistics, diagnostic testing and the consequences of violations of the CLRM assumptions, along with plausible remedial steps. Model-building philosophies are discussed with particular reference to the general-to-specific approach. Finally, the main principles of principal components analysis are briefly discussed in an appendix. Applications covered in this chapter include hedonic models of rental values and 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. How forecasts can be generated from such models is discussed, and upon what criteria these forecasts can be evaluated. Examples include model-building for stock returns and dividends, 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 chapter also focuses on how such models are estimated, and how restrictions are tested and imposed. 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 univariate 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 (ppp) exchange rates hypothesis.

Chapter 8

This covers the highly popular 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 changes, 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 presents an introduction to what is arguably one of the most rapidly developing areas in financial modelling: that of simulations. 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 11

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 12

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.