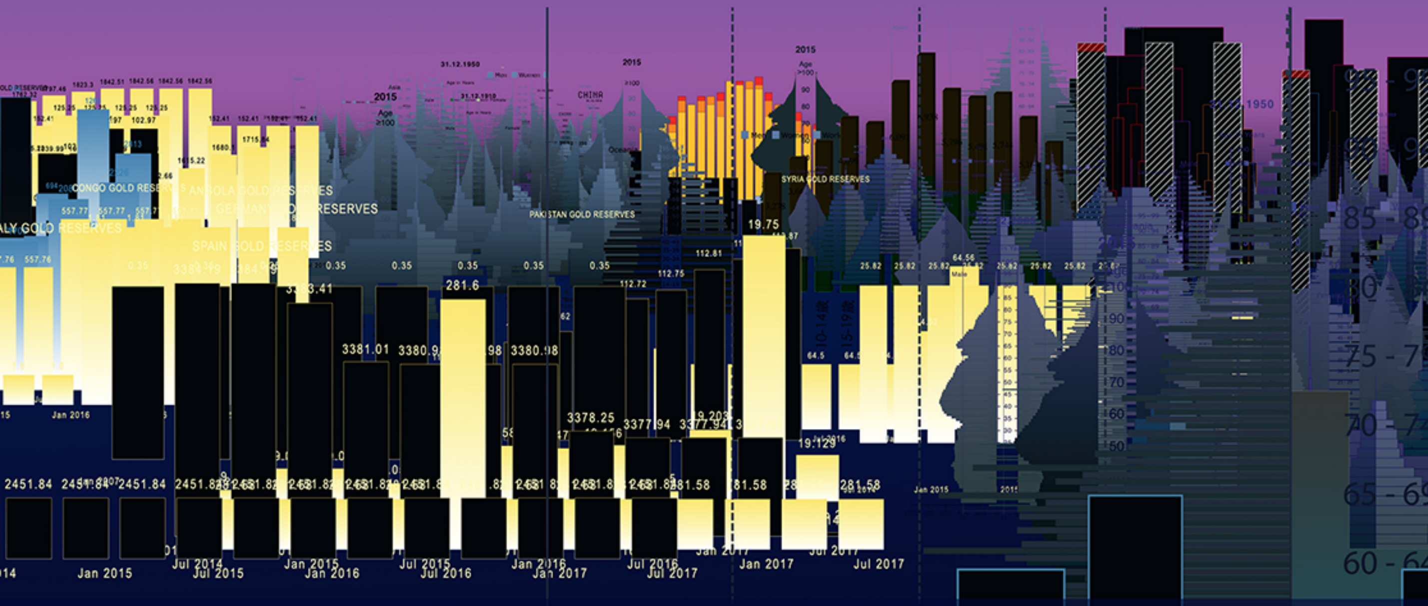
CHRIS BROOKS



INTRODUCTORY ECONOMETRICS FOR FINANCE

4TH EDITION

Introductory Econometrics for Finance

This bestselling and thoroughly classroom-tested textbook is a complete resource for finance students. A comprehensive and illustrated discussion of the most common empirical approaches in finance prepares students for using econometrics in practice, while detailed case studies help them understand how the techniques are used in relevant financial contexts. Learning outcomes, key concepts and end-of-chapter review questions (with full solutions online) highlight the main chapter takeaways and allow students to self-assess their understanding. Building on the successful data- and problem-driven approach of previous editions, this fourth edition has been updated with new examples, additional introductory material on mathematics and dealing with data, as well as more advanced material on extreme value theory, the generalised method of moments and state space models. A dedicated website, with numerous student and instructor resources including videos and a set of companion manuals for various statistical software – all available free of charge – completes the learning package.

Chris Brooks is Professor of Finance at the ICMA Centre, Henley Business School, University of Reading, UK where he also obtained his PhD. Chris has diverse research interests and has published over a hundred articles in leading academic and practitioner journals, and six books. He is Associate Editor of several journals, including the *Journal of Business Finance and Accounting* and the *British Accounting Review*. He acts as consultant and advisor for various banks, corporations and regulatory and professional bodies in the fields of finance, real estate and econometrics.

Introductory Econometrics for Finance

FOURTH EDITION

CHRIS BROOKS
The ICMA Centre, Henley Business School, University of Reading

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108422536
DOI: 10.1017/9781108524872

First published 2002
Second edition published 2008
Third edition published 2014
Fourth edition published 2019

Printed and bound in Great Britain by Clays Ltd, Elcograf S.p.A.

A catalogue record for this publication is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Names: Brooks, Chris, 1971– author.
Title: Introductory econometrics for finance / Chris Brooks, The ICMA Centre, Henley Business School, University of Reading.
Description: Fourth edition. | Cambridge, United Kingdom ; New York, NY : Cambridge University Press, 2019. | Includes bibliographical references and index.
Identifiers: LCCN 2018061692 | ISBN 9781108422536 (hardback : alk. paper) | ISBN 9781108436823 (pbk. : alk. paper)
Subjects: LCSH: Finance–Econometric models. | Econometrics.
Classification: LCC HG173 .B76 2019 | DDC 332.01/5195–dc23
LC record available at https://lccn.loc.gov/2018061692

ISBN 978-1-108-42253-6 Hardback
ISBN 978-1-108-43682-3 Paperback

Additional resources for this publication at www.cambridge.org/brooks4

Contents in Brief

Detailed Contents

Chapter 10 Switching and State Space Models 447

Chapter 11 Panel Data 490

Chapter 12 Limited Dependent Variable Models 516

Figures

Tables

Boxes

Screenshots

Preface to the Fourth Edition

All of the motivations for the first edition, described below, seem just as important today. Given that the book seems to have gone down well with readers, I have left the style largely unaltered but added a lot of new material. The main motivations for writing the first edition of the book were:

- To write a book that focused on *using and applying* the techniques rather than deriving proofs and learning formulae.
- To write an accessible textbook that required no prior knowledge of econometrics, but which also covered more recently developed approaches usually only found in more advanced texts.
- To use examples and terminology from finance rather than economics since there are many introductory texts in econometrics aimed at students of economics but none for students of finance.
- To populate the book with case studies of the use of econometrics in practice taken from the academic finance literature.
- To include sample instructions, screen dumps and computer output from a popular econometrics package. This enabled readers to see how the techniques can be implemented in practice. In this fourth edition, the EViews instructions have been separated off and are available free of charge on the book's web site along with parallel manuals for other packages including Stata, Python and R.
- To develop a companion web site containing answers to end of chapter questions, a multiple choice question bank with feedback, PowerPoint slides and other supporting materials.

What is New in the Fourth Edition

The fourth edition includes a number of important new features

(1) Students of finance have enormously varying backgrounds, and in particular varying levels of training in elementary mathematics and statistics. In order to make the book more self-contained, the introductory chapter has again been expanded. So the material previously in Chapter 2 has been separated into introductory maths (Chapter 1) and introductory statistics/dealing with data (Chapter 2).

(2) More new material has been added on state space models and their estimation using the Kalman filter in Chapter 10.

(3) A chapter has been added which collects together a number of techniques often used in financial research, including event studies and the Fama MacBeth approach (previously elsewhere in the book) and new sections on using extreme value distribution to model the fat tails in financial series and on estimating models with the generalised method of moments.

(4) The incorporation of EViews directly into the core of the book may have been a distraction for those using other packages. Thus, as stated above, in the new edition the EViews instructions have been separated off and are available free of charge on the book's web site along with parallel manuals for other packages including Stata, Python and R. This package should ensure that the book fits the bill whatever the reader's preferred software.

Motivations for the First Edition

This book had its genesis in two sets of lectures given annually by the author at the ICMA Centre (formerly the ISMA Centre), Henley Business School, University of Reading and arose partly from several years of frustration at the lack of an appropriate textbook. In the past, finance was but a small sub-discipline drawn from economics and accounting, and therefore it was generally safe to assume that students of finance were well grounded in economic principles; econometrics would be taught using economic motivations and examples.

However, finance as a subject has taken on a life of its own in recent years. Drawn in by perceptions of exciting careers in the financial markets, the number of students of finance has grown phenomenally all around the world. At the same time, the diversity of educational backgrounds of students taking finance courses has also expanded. It is not uncommon to find undergraduate students of finance even without advanced high-school qualifications in mathematics or economics. Conversely, many with PhDs in physics or engineering are also attracted to study finance at the Masters level. Unfortunately, authors of textbooks failed to keep pace with the change in the nature of students. In my opinion, the currently available textbooks fall short of the requirements of this market in three main regards, which this book seeks to address

(1) Books fall into two distinct and non-overlapping categories: the introductory and the advanced. Introductory textbooks are at the appropriate level for students with limited backgrounds in mathematics or statistics, but their focus is too narrow. They often spend too long deriving the most basic results, and treatment of important, interesting and relevant topics (such as simulations methods, VAR modelling, etc.) is covered in only the last few pages, if at all. The more advanced textbooks, meanwhile, usually require a quantum leap in the level of mathematical ability assumed of readers, so that such books cannot be used on courses lasting only one or two semesters, or where students have differing backgrounds. In this book, I have tried to sweep a broad brush over a large number of different econometric techniques that are relevant to the analysis of financial and other data.

(2) Many of the currently available textbooks with broad coverage are too theoretical in nature and students can often, after reading such a book, still have no idea of how to tackle real-world problems themselves, even if they have mastered the techniques in theory. This book and the accompanying software manuals should assist students who wish to learn how to estimate models for themselves – for example, if they are required to complete a project or dissertation. Some examples have been developed especially for this book, while many others are drawn from the academic finance literature. In my opinion, this is an essential but rare feature of a textbook that should help to show students how econometrics is really applied. It is also hoped that this approach will encourage some students to delve deeper into the literature, and will give useful pointers and stimulate ideas for research projects. It should, however, be stated at the outset that the purpose of including examples from the academic finance print is not to provide a comprehensive overview of the literature or to discuss all of the relevant work in those areas, but rather to illustrate the techniques. Therefore, the literature reviews may be considered deliberately deficient, with interested readers directed to the suggested readings and the references therein.

(3) With few exceptions, almost all textbooks that are aimed at the introductory level draw their motivations and examples from economics, which may be of limited interest to students of finance or business. To see this, try motivating regression relationships using an example such as the effect of changes in income on consumption and watch your audience, who are primarily interested in business and finance applications, slip away and lose interest in the first ten minutes of your course.

Who Should Read this Book?

The intended audience is undergraduates or Masters/MBA and PhD students who require a broad knowledge of modern econometric techniques commonly employed in the finance literature. It is hoped that the book will also be useful for researchers (both academics and practitioners), who require an introduction to the statistical tools commonly employed in the area of finance. The book can be used for courses covering financial time-series analysis or financial econometrics in undergraduate or post-graduate programmes in finance, financial economics, securities and investments.

Although the applications and motivations for model-building given in the book are drawn from finance, the empirical testing of theories in many other disciplines, such as management studies, business studies, real estate, economics and so on, may usefully employ econometric analysis. For this group, the book may also prove useful.

Finally, while the present text is designed mainly for students at the undergraduate or Masters level, it could also provide introductory reading in financial modelling for finance doctoral programmes where students have backgrounds which do not include courses in modern econometric techniques.

Pre-Requisites for Good Understanding of This Material

In order to make the book as accessible as possible, no prior knowledge of statistics, econometrics or algebra is required, although those with a prior exposure to calculus, algebra (including matrices) and basic statistics will be able to progress more quickly. The emphasis throughout the book is on a valid application of the techniques to real data and problems in finance.

In the finance and investment area, it is assumed that the reader has knowledge of the fundamentals of corporate finance, financial markets and investment. Therefore, subjects such as portfolio theory, the capital asset pricing model (CAPM) and arbitrage pricing theory (APT), the efficient markets hypothesis, the pricing of derivative securities and the term structure of interest rates, which are frequently referred to throughout the book, are not explained from first principles in this text. There are very many good books available in corporate finance, in investments and in futures and options, including those by Brealey and Myers (2013), Bodie, Kane and Marcus (2014) and Hull (2017) respectively.

Acknowledgements

I am grateful to Gita Persand, Olan Henry, James Chong and Apostolos Katsaris, who assisted with various parts of the software applications for the first edition. I am also grateful to Hilary Feltham for assistance with Chapters 1 and 2. I would also like to thank Simon Burke, James Chong and Con Keating for detailed and constructive comments on various drafts of the first edition, Simon Burke for suggestions on parts of the second edition, Mike Clements, Jo Cox, Eunyoung Mallet, Ogonna Nneji, Ioannis Oikonomou and Chardan Wese Simen for comments on part of the third edition and Marcel Prokopczuk for comments on part of the fourth edition.

I have additionally benefited from the comments, suggestions and questions of the following list of people, many of whom sent useful e-mails pointing out typos or inaccuracies: Zary Aftab, Panos Ballis-Papanastasiou, Mirco Balatti, Peter Burridge, Kyongwook Choi, Rishi Chopra, Araceli Ortega Diaz, Xiaoming Ding, Thomas Eilertsen, Waleid Eldien, Junjong Eo, Merlyn Foo, Andrea Gheno, Christopher Gilbert, Kimon Gomozias, Jan de Gooijer and his colleagues, Cherif Guermat, Abid Hameed, Ibrahim Jamali, Kejia Jia, Arty Khemlani, Margaret Lynch, David McCaffrey, Tehri Jokipii, Emese Lazar, Zhao Liuyan, Dimitri Lvov, Bill McCabe, Junshi Ma, Raffaele Mancuso, David Merchan, Yue Min, Victor Murinde, Kyoung Gook Park, Mikael Petitjean, Marcelo Perlin, Thai Pham, Jean-Sebastien Pourchet, Marcel Prokopczuk, Tao Qingmei, Satya Sahoo, Lisa Schopohl, Guilherme Silva, Jerry Sin, Andre-Tudor Stancu, Silvia Stanescu, Fred Sterbenz, Birgit Strikholm, Yiguo Sun, Li Qui, Panagiotis Varlagas, Jakub Vojtek, Henk von Eije, Jue Wang, Robert Wichmann and Meng-Feng Yen.

Outline of the Remainder of this Book

Chapter 1

This covers the key mathematical techniques that readers will need some familiarity with to be able to get the most out of the remainder of this book. It starts with a discussion of what econometrics is about and how to set up an econometric model, then moves on to present the mathematical material on functions, and powers, exponents and logarithms of numbers. It then proceeds to explain the basics of differentiation and matrix algebra, which is illustrated via the construction of optimal portfolio weights.

Chapter 2

This chapter presents the statistical foundations of econometrics and the beginnings of how to work with financial data. It covers key results in statistics, discusses probability distributions, how to summarise data and different types of data. The chapter then moves on to discuss the calculation of present and future values, compounding and discounting, and how to calculate nominal and real returns in various ways.

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. 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 4

This continues and develops the material of Chapter 3 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 (PCA) to interest rates.

Chapter 5

Chapter 5 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 6

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 7

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 8

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 8 include spot and futures markets, tests for cointegration between international bond markets and tests of the purchasing power parity (PPP) hypothesis and of the expectations hypothesis of the term structure of interest rates.

Chapter 9

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 and conditional correlation models are described, and applications to the estimation of conditional betas and time-varying hedge ratios, and to financial risk measurement, are given.

Chapter 10

This begins by discussing how to test for and model 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 then 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. Finally, the second part of the chapter moves on to examine how to specify models with time-varying parameters using the state space form and how to estimate them with the Kalman filter.

Chapter 11

This 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 12

This 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 13

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 14

This chapter presents a collection of techniques that are particularly useful for conducting research in finance. It begins with detailed illustrations of how to conduct event studies, which are commonly used in corporate finance applications, and how to use the Fama-French factor model approach to asset pricing. The chapter then proceeds to present the families of extreme value models that are used to accurately capture the fat tails of asset return distributions and as the basis for value at risk calculations. Finally, the chapter covers the generalised method of moments (GMM) technique, which has become increasingly popular in recent years for estimating a range of different types of models in finance.

Chapter 15

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.

1

Introduction and Mathematical Foundations

LEARNING OUTCOMES

In this chapter, you will learn how to

- Describe the key steps involved in building an econometric model
- Work with powers, exponents and logarithms
- Plot, interpret and calculate the roots of functions
- Use sigma (Σ) and pi (Π) notation
- Apply rules to differentiate various types functions
- Work with matrices
- Calculate the trace, inverse and eigenvalues of a matrix
- Construct and interpret utility functions

Learning econometrics is in many ways like learning a new language. To begin with, nothing makes sense and it is as if it is impossible to see through the fog created by all the unfamiliar terminology. While the way of writing the models – the *notation* – may make the situation appear more complex, in fact it is supposed to achieve the exact opposite. The ideas themselves are mostly not so complicated, it is just a matter of learning enough of the language that everything fits into place. So if you have never studied the subject before, then persevere through this preliminary chapter and you will hopefully be on your way to being fully fluent in econometrics!

This chapter comprises two parts. The first sets the scene for the book by discussing in broad terms the questions of what econometrics is, and the kinds of problems that can be tackled using econometrics. The second part of the chapter covers the mathematical techniques that underpin approaches to modelling and dealing with data in finance. Those with some prior background in algebra and introductory mathematics may skip the second part of this chapter without loss of continuity, but hopefully the material will also constitute a useful refresher for those who have studied mathematics but a long time ago!

1.1 What is Econometrics?

The literal meaning of the word 'econometrics' is 'measurement in economics'. The first five 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 Examples of the uses 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'?

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 sometimes 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 usually of less 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 or 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.

One of the most rapidly evolving areas of financial application of statistical tools is in the modelling of market microstructure problems. ‘Market microstructure’ may broadly be defined as the process whereby *investors’ preferences and desires are translated into financial market transactions*. It is evident that microstructure effects are important and represent a key difference between financial and other types of data. These effects can potentially impact on many other areas of finance. For example, market rigidities or frictions can imply that current asset prices do not fully reflect future expected cashflows (see the discussion in Chapter 10 of this book). Also, investors are likely to require compensation for holding securities that are illiquid, and therefore embody a risk that they will be difficult to sell owing to the relatively high probability of a lack of willing purchasers at the time of desired sale. Measures such as volume or the time between trades are sometimes used as proxies for market liquidity.

A comprehensive survey of the literature on market microstructure is given by Madhavan (2000). He identifies several aspects of the market microstructure literature, including price formation and price discovery, issues relating to market structure and design, information and disclosure. There are also relevant books by O’Hara (1995), Harris (2002) and Hasbrouck (2007). At the same time, there has been considerable advancement in the sophistication of econometric models applied to microstructure problems. For example, an important innovation was the autoregressive conditional duration (ACD) model attributed to Engle and Russell (1998). An interesting application can be found in Dufour and Engle (2000), who examine the effect of the time between trades on the price-impact of the trade and the speed of price adjustment.

1.3 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.

- *Steps 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.

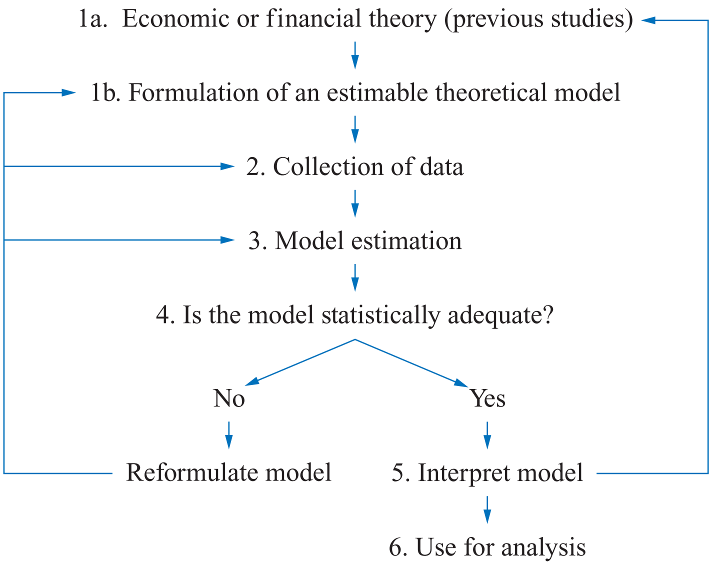


Figure 1.1 Steps involved in formulating an econometric model

- *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 the 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.

1.4 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 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

BOX 1.2 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?

journal. The questions that are always worth asking oneself when reading a paper are outlined in Box 1.2.

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.

This chapter now moves on to cover the fundamental mathematical framework that underpins financial econometrics. This material is intended as a refresher for readers who have covered these topics in the past but require a reminder; students who are seeing these concepts for the first time may find a more thorough treatment covering an entire book useful in addition to this text – see, for example Renshaw (2016) or Swift and Piff (2014), which are both detailed and very accessible.

1.5 Functions

1.5.1 Introduction to Functions

The ultimate objective of econometrics is usually to build a model, which may be thought of as a simplified version of the true relationship between two or more variables that can be described by a *function*. A function is simply a mapping or relationship between an input or set of inputs and an output. We usually write that y, the output, is a function f of x, the input, so $y = f(x)$. $f(.)$ is simply a general method of stating that y is related to x in some fashion. Another way to say this is that f provides a mapping between y and x so that it tells us, for every given value of x, what the corresponding value of y would be. f is a unique (1:1) mapping so that for each value of x there is only one corresponding value of y.

The *domain* of x is defined as the set of values that this variable can take; the *range* refers to the respective set of values that y can take. Usually, neither the domain nor the range are specified, in which case they can both be assumed to be allowed to take any real values.

1.5.2 Straight Lines

y could be a linear function of x, where the relationship can be expressed as a straight line on a graph, or y could be a non-linear function of x, in which case the relationship between the two variables would be represented graphically as a curve. If the relationship is linear, we could write the equation for this straight line as

$$y = a + bx \tag{1.1}$$

y and x are called *variables*, while a and b are *parameters*; a is termed the *intercept* and b is the *slope* or *gradient* of the line. The intercept is the point at which the line crosses the y-axis, while the slope measures the steepness of the line. Note that there will be only one value of a and one value of b, although there will be many values of x and of y. a and b could each be any combination of positive, negative or zero.

Table 1.1 Sample data on hours of study and grades

Hours of study (x)	Grade-point average in % (y)
0	25
100	30
400	45
800	65
1000	75
1200	85

To illustrate, suppose we were trying to model the relationship between a student's grade-point average y (expressed as a percentage), and the number of hours that they studied throughout the year, x. Suppose further that the relationship can be written as a linear function with $y = 25 + 0.05x$.

Clearly it is unrealistic to assume that the link between grades and hours of study follows a straight line, but let us keep this assumption for now. So the intercept of the line, a, is 25, and the slope, b, is 0.05. What does this equation mean? It means that a student spending no time studying at all ($x = 0$) could expect to earn a 25% average grade, and for every hour of study time, their average grade should improve by 0.05% – in other words, an extra 100 hours of study through the year would lead to a 5% increase in the grade.

Suppose that a particular student wished to score a perfect 100% grade-point average. How many hours would (s)he need to study? To calculate this, we would need to set $y = 100$ and then to solve for x: $100 = 25 + 0.05x$, so $x = 1500$ hours. We could construct a table with several values of x and the corresponding value of y as in Table 1.1 and then plot them onto a graph (Figure 1.2).

We can see from the graph that the gradient of this line is positive (i.e., it slopes upwards from left to right). Note that for a straight line, the slope is the same along the whole line; this slope can be calculated from a graph by taking any two points on the line and dividing the change in the value of y by the change in the value of x between the two points.

In general, a capital delta, Δ, is used to denote a change in a variable. For example, suppose that we want to take the two points $x = 100$, $y = 30$ and $x = 1000$, $y = 75$. We could write these two points using a coordinate notation (x,y) and so (100,30) and (1000,75) in this example. We would calculate the slope of the line as

$$\frac{\Delta y}{\Delta x} = \frac{75 - 30}{1000 - 100} = 0.05 \tag{1.2}$$

So indeed, we have confirmed that the slope is 0.05 (although in this case we knew that from the start). Two other examples of straight line graphs are given in Figure 1.3.

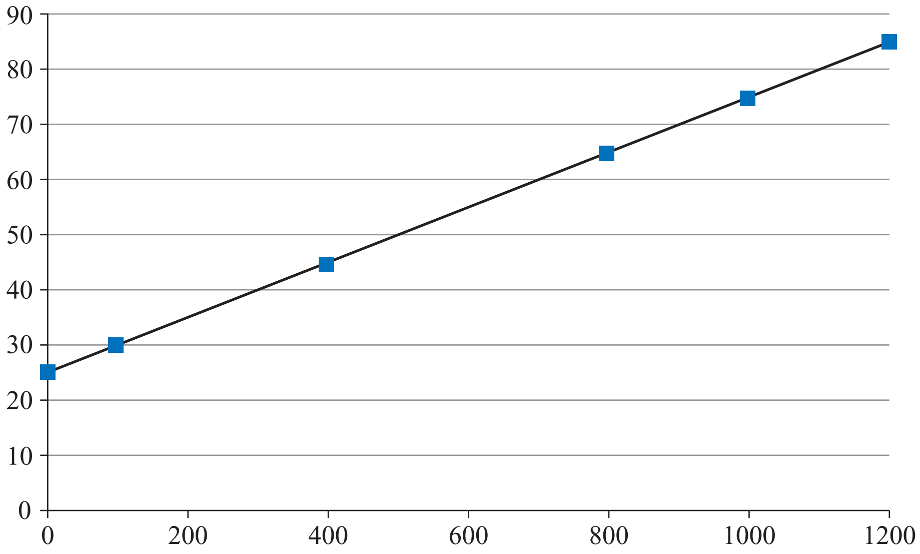


Figure 1.2 A plot of hours studied (x) against grade-point average (y)

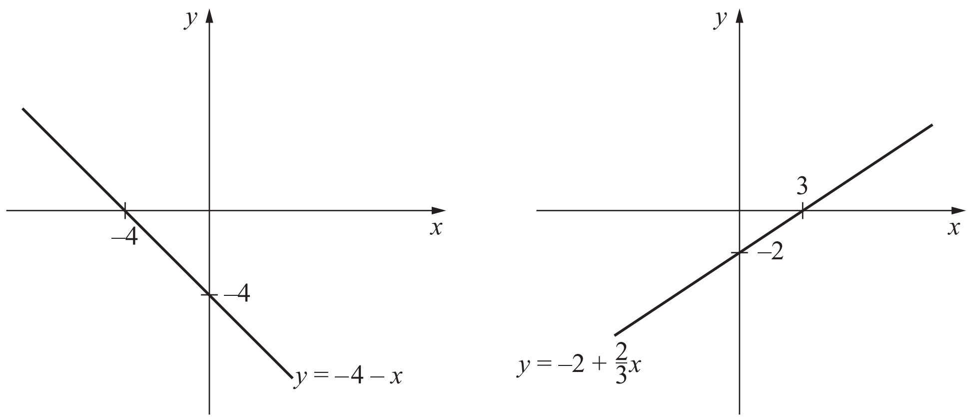


Figure 1.3 Examples of different straight line graphs

The gradient of the line can be zero or negative instead of positive. If the gradient is zero, the resulting plot will be a flat (horizontal) straight line. We could then write it as $y = 25 + 0x$, so that whatever the value of x, y will always be the same (25).

If there is a specific change in x, Δx, and we want to calculate the corresponding change in y, we would simply multiply the change in x by the slope, so $\Delta y = b\Delta x$.

As a final point, note that we stated above that the point at which a function crosses the y-axis is termed the intercept. The point at which the function crosses the x-axis is called its *root*. In the example above, if we take the function $y = 25 + 0.05x$, set y to zero and rearrange the equation, we would find that the root would be $x = -500$.

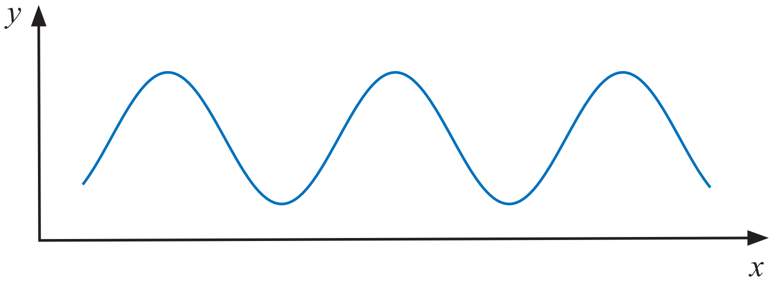


Figure 1.4 Example of a general polynomial function

In this case, the root of the equation does not have a useful interpretation (as the number of hours studied cannot be negative) but this will not always be the case.

The equation for a straight line has one root (except for a horizontal straight line such as $y = 4$, where there would be no root since it never crosses the x-axis). Further examples of how to calculate the roots of an equation will be given in Section 1.5.3.

1.5.3 Polynomial Functions

A linear function is often not sufficiently flexible to be able to accurately describe the relationship between two variables, and so a quadratic function may be used instead. A *polynomial* simply adds higher order powers of the variable x into the function. In the most general case, we would have an n^{th} order polynomial (a polynomial of order n)

$$y = a + b_1 x + b_2 x^2 + b_3 x^3 + \cdots + b_n x^n \tag{1.3}$$

If $n = 2$, we have a quadratic equation, if $n = 3$ a cubic, if $n = 4$ a quartic and so on. We use polynomials if y depends only on one variable x but in a non-linear way (and so it cannot be expressed as a straight line). An example of the shape of a general polynomial function is given in Figure 1.4.

Broadly, the higher the order of the polynomial, the more complex will be the relationship between y and x and the more twists and turns there will be in the plot like Figure 1.4. However, usually $n = 2$, a quadratic equation, is sufficient to describe the function as it seems unlikely that a real series y will rise with x then fall before rising again and so on, which would be the case if it was described by a higher order polynomial. So now we will focus on the quadratic case.

We could write the general expression for a quadratic function as

$$y = a + bx + cx^2 \tag{1.4}$$

where x and y are again the variables and a, b, c are the parameters that describe the shape of the function. Note that we have changed notation slightly for simplicity between equations (1.3) and (1.4), writing the slope parameters as b and c rather than

BOX 1.3 The roots of a quadratic equation

- A quadratic equation has two roots
- The roots may be distinct (i.e., different from one another), or they may be the same (repeated roots); they may be real numbers (e.g., 1.7, −2.357, 4, etc.) or what are known as *complex numbers*
- The roots can be obtained either by *factorising* the equation – i.e., contracting it into parentheses, by 'completing the square' or by using the formula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2c} \quad (1.5)$$

- If $b^2 > 4ac$, the function will have two unique roots and it will cross the x-axis in two separate places; if $b^2 = 4ac$, the function will have two equal roots and it will only cross the x-axis in one place; if $b^2 < 4ac$, the function will have no real roots (only complex roots), it will not cross the x-axis at all and thus the function will always be above the x-axis.

b_1 and b_2. Either notation is equally acceptable so long as we are clear and explain what we mean.

A linear function only has two parameters (the intercept, a and the slope, b), but a quadratic has three and hence it is able to adapt to a broader range of relationships between y and x. The linear function is a special case of the quadratic where c is zero. As before, a is the intercept and defines where the function crosses the y-axis; the parameters b and c determine the shape.

Quadratic equations can be either ∪-shaped or ∩-shaped. As x becomes very large and positive or very large and negative, the x^2 term will dominate the behaviour of y and it is thus c that determines which of these shapes will apply. Figure 1.5 shows two examples of quadratic functions – in the first case c is positive and so the curve is ∪-shaped, while in the second c is negative so the curve is ∩-shaped. We discussed above that the root(s) of an equation is (are) the place(s) where the line crosses the x-axis. Box 1.3 discusses the features of the roots of a quadratic equation and shows how to calculate them.

EXAMPLE 1.1

Determine the roots of the following quadratic equations

1. $y = x^2 + x - 6$
2. $y = 9x^2 + 6x + 1$
3. $y = x^2 - 3x + 1$
4. $y = x^2 - 4x$

SOLUTION We would solve these equations by setting them in turn to zero. We could then use the quadratic formula from equation (1.5) in each case, although it is usually quicker to determine first whether they factorise (see Box 1.3).

1. $x^2 + x - 6 = 0$ factorises to $(x - 2)(x + 3) = 0$ and thus the roots are 2 and -3, which are the values of x that set the function to zero. In other words, the function will cross the x-axis at $x = 2$ and $x = -3$.

2. $9x^2 + 6x + 1 = 0$ factorises to $(3x + 1)(3x + 1) = 0$ and thus the roots are $-\frac{1}{3}$ and $-\frac{1}{3}$. This is known as repeated roots – since this is a quadratic equation there will always be two roots but in this case they are both the same. We call the expression $9x^2 + 6x + 1$ a *perfect square*. Here the plot of y against x would touch, but not cross, the x-axis at $x = -\frac{1}{3}$.

3. $x^2 - 3x + 1 = 0$ does not factorise and so the formula must be used with $a = 1, b = -3, c = 1$ and the roots are 0.38 and 2.62 to two decimal places.

4. $x^2 - 4x = 0$ factorises to $x(x - 4) = 0$ and so the roots are 0 and 4. The function crosses the x-axis at the points (0,0) and (4,0).

Note that all of these equations have two real roots. If we had an equation such as $y = 3x^2 - 2x + 4$, this would not factorise and would have complex roots since $b^2 - 4ac < 0$ in the quadratic formula. A similar situation is illustrated in the left-hand part of Figure 1.5, which does not cross the x-axis anywhere.

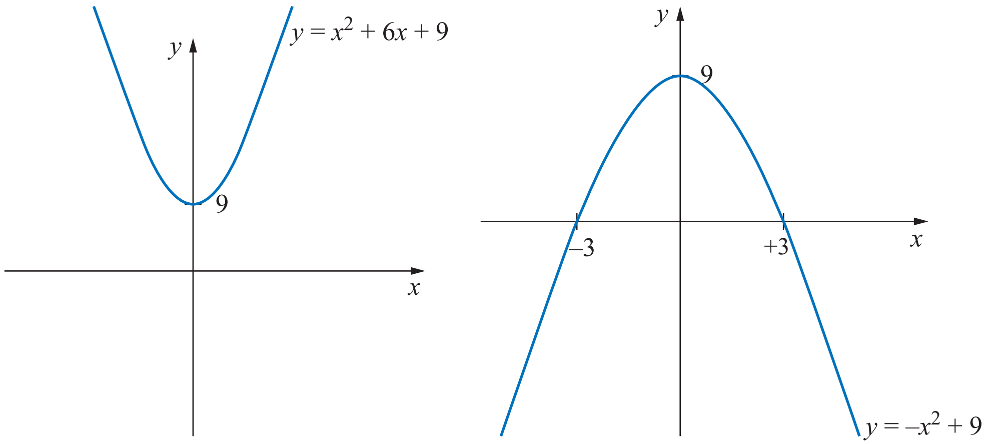


Figure 1.5 Examples of quadratic functions

1.5.4 Powers of Numbers or of Variables

A number or variable raised to a power is simply a way of writing repeated multiplication. So, for example, raising x to the power 2 means squaring it (i.e., $x^2 = x \times x$); raising it to the power 3 means cubing it ($x^3 = x \times x \times x$), and so on. The number that we are raising the number or variable to is called the *index*, so for x^3, 3 would be the index. There are a few rules for manipulating powers and their indices given in Box 1.4.

1.5.5 The Exponential Function

It is sometimes the case that the relationship between two variables is best described by an *exponential* function – for example, when a variable y grows (or reduces) at

BOX 1.4 Manipulating powers and their indices

- Any number or variable raised to the power one is simply that number or variable, e.g., $3^1 = 3$, $x^1 = x$, and so on.
- Any number or variable raised to the power zero is one, e.g., $5^0 = 1$, $x^0 = 1$, etc., except that 0^0 is not defined (i.e., it does not exist).
- If the index is a negative number, this means that we divide one by that number – for example, $x^{-3} = \frac{1}{x^3} = \frac{1}{x \times x \times x}$.
- If we want to multiply together a given number raised to more than one power, we would add the corresponding indices together – for example, $x^2 \times x^3 = x^2 x^3 = x^{2+3} = x^5$. The general rule is $x^a \times x^b = x^{a+b}$.
- If we want to calculate the power of a variable raised to a power (i.e., the power of a power), we would multiply the indices together – for example, $(x^2)^3 = x^{2 \times 3} = x^6$. The general rule is $(x^a)^b = x^{a \times b}$.
- If we want to divide a variable raised to a power by the same variable raised to another power, we subtract the second index from the first – for example, $\frac{x^3}{x^2} = x^{3-2} = x$. The general rule is $\frac{x^a}{x^b} = x^{a-b}$.
- If we want to divide a variable raised to a power by a different variable raised to the same power, the following result applies

$$\left(\frac{x}{y}\right)^n = \frac{x^n}{y^n}.$$

- The power of a product is equal to each component raised to that power – for example, $(x \times y)^3 = x^3 \times y^3$.
- It is important to note that the indices for powers do not have to be integers. For example, $x^{\frac{1}{2}}$ is the notation we would use for taking the square root of x, sometimes written $\sqrt{x}$. Other, non-integer powers are also possible, but are harder to calculate by hand (e.g., $x^{0.76}$, $x^{-0.27}$, etc.) In general, $x^{1/n} = \sqrt[n]{x}$, the nth root of x.

a rate in proportion to its current value x, in which case we would write $y = e^x$. e is a simply number: 2.71828...In fact, e can be derived by letting n in the following expression tend towards infinity

$$e \approx \left(1 + \frac{1}{n}\right)^n \tag{1.6}$$

Alternatively, we can define e as the result from the following infinite sum

$$e = \sum_{i=0}^{\infty} \frac{1}{i!} = \frac{1}{1} + \frac{1}{1} + \frac{1}{6} + \frac{1}{24} + \ldots \tag{1.7}$$

where ! denotes a factorial (e.g., $4! = 4 \times 3 \times 2 \times 1$).

The exponential function has several useful properties, including that it is its own derivative (see Section 1.6.1 below) and thus the gradient of the function e^x at any point is also e^x; it is also useful for capturing the increase in value of an amount of money that is subject to compound interest. The exponential function can never be negative, so when x is negative, y is close to zero but positive. It crosses the y-axis at one and the slope increases at an increasing rate from left to right, as shown in Figure 1.6.

1.5.6 Logarithms

Logarithms were invented before computers and pocket calculators were widely available to simplify cumbersome calculations, since exponents can then be added or subtracted, which is easier than multiplying or dividing the original numbers. While logarithmic transformations are no longer necessary for computational ease, they still have important uses in algebra and in data analysis. For the latter, there are at least three reasons why log transforms may be useful. First, taking a logarithm

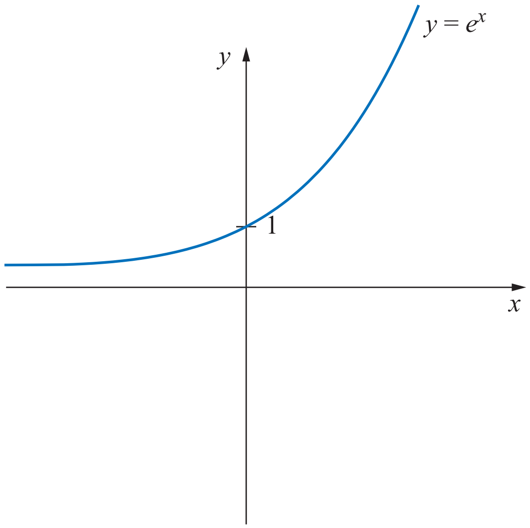


Figure 1.6 A plot of an exponential function

BOX 1.5 The laws of logs

For variables x and y

- $\ln(xy) = \ln(x) + \ln(y)$
- $\ln(x/y) = \ln(x) - \ln(y)$
- $\ln(y^c) = c\ln(y)$
- $\ln(1) = 0$
- $\ln(1/y) = \ln(1) - \ln(y) = -\ln(y)$.
- $\ln(e^x) = e^{\ln(x)} = x$

can often help to rescale the data so that their variance is more constant, which overcomes a common statistical problem known as *heteroscedasticity*, discussed in detail in Chapter 5. Second, logarithmic transforms can help to make a positively skewed distribution closer to a normal distribution. Third, taking logarithms can also be a way to make a non-linear, multiplicative relationship between variables into a linear, additive one. These issues will also be discussed in some detail in Chapter 5.

To motivate how logs work, consider the power relationship $2^3 = 8$. Using logarithms, we would write this as $\log_2 8 = 3$, or 'the log to the base 2 of 8 is 3'. Hence we could say that a logarithm is defined as the power to which the base must be raised to obtain the given number. More generally, if $a^b = c$, then we can also write $\log_a c = b$.

Natural logarithms, also known as logs to base e, are more commonly used and more useful mathematically than logs to any other base. A log to base e is known as a *natural* or *Napierian* logarithm, denoted interchangeably by $\ln(y)$ or $\log(y)$. Taking a natural logarithm is the inverse of a taking an exponential, so sometimes the exponential function is called the *antilog*. The log of a number less than one will be negative, e.g., $\ln(0.5) \approx -0.69$. We cannot take the log of a negative number (so $\ln(-0.6)$, for example, does not exist). The properties of logarithmic functions or 'laws of logs' describe the way that we can work with logs or manipulate expressions using them. These are presented in Box 1.5.

If we plot a log function, $y = \ln(x)$, it would cross the x-axis at one, as in Figure 1.7. It can be seen that as x increases, y increases at a slower rate, which is the opposite to an exponential function where y increases at a faster rate as x increases.

1.5.7 Inverse Functions

If we have a function such that $y = f(x)$, we would write the inverse as $x = f^{-1}(y)$. To give a simple example of a linear equation, if $y = 6x - 3$, the inverse function would be a rearrangement of the function to make x the subject: $x = (y + 3)/6$. For polynomials of order n, there could be up to n possible inverse functions, although the inverse of a function will not always exist.

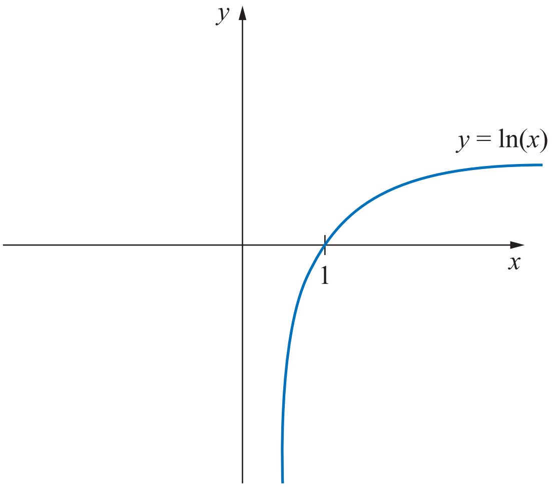


Figure 1.7 A plot of a logarithmic function

1.5.8 Sigma Notation

If we wish to add together several numbers (or observations from variables), the *sigma* or summation operator can be very useful. Σ means 'add up all of the following elements'. For example, $\Sigma(1, 2, 3) = 1 + 2 + 3 = 6$. In the context of adding the observations on a variable, it is helpful to add 'limits' to the summation (although note that the limits are not always written out if the meaning is obvious without them). So, for instance, we might write

$$\sum_{i=1}^{4} x_i$$

where the i subscript is called an index, 1 is the lower limit and 4 is the upper limit of the sum. This would mean adding all of the values of x from x_1 to x_4.

It might be the case that one or both of the limits is not a specific number – for instance, $\sum_{i=1}^{n} x_i$, which would mean $x_1 + x_2 + \ldots + x_n$, or sometimes we simply write $\sum_i x_i$ to denote a sum over all the values of the index i. It is also possible to construct a sum of a more complex combination of variables, such as $\sum_{i=1}^{n} x_i z_i$, where x_i and z_i are two separate random variables.

It is important to be aware of a few properties of the sigma operator. For example, the sum of the observations on a variable x plus the sum of the observations on another variable z is equivalent to the sum of the observations on x and z first added together individually

$$\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} z_i = \sum_{i=1}^{n} (x_i + z_i) \tag{1.8}$$

The sum of the observations on a variable x each multiplied by a constant c is equivalent to the constant multiplied by the sum

$$\sum_{i=1}^{n} cx_i = c\sum_{i=1}^{n} x_i \tag{1.9}$$

But the sum of the products of two variables is not the same as the product of the sums

$$\sum_{i=1}^{n} x_i z_i \neq \sum_{i=1}^{n} x_i \sum_{i=1}^{n} z_i \tag{1.10}$$

We can write the left-hand side (LHS) of equation (1.10) as

$$\sum_{i=1}^{n} x_i z_i = x_1 z_1 + x_2 z_2 + \ldots + x_n z_n \tag{1.11}$$

whereas the right-hand side (RHS) of equation (1.10) is written

$$\sum_{i=1}^{n} x_i \sum_{i=1}^{n} z_i = (x_1 + x_2 + \ldots + x_n)(z_1 + z_2 + \ldots + z_n) \tag{1.12}$$

We can see that equations (1.11) and (1.12) are different since the latter contains many 'cross-product' terms such as $x_1 z_2$, $x_3 z_6$, $x_9 z_2$, etc., whereas the former does not.

If we sum n identical elements (i.e., we add a given number to itself n times), we obtain n times that number

$$\sum_{i=1}^{n} x = x + x + \ldots + x = nx \tag{1.13}$$

Suppose that we sum all of the n observations on a series, x_i – for example, the x_i could be the daily returns on a stock (which are not all the same), we would obtain

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \ldots + x_n = n\bar{x} \tag{1.14}$$

So the sum of all of the observations is, from the definition of the mean, equal to the number of observations multiplied by the mean of the series, $\bar{x}$. Notice that the difference between this situation in equation (1.14) and the previous one in equation (1.13) is that now the x_i are different from one another whereas before they were all the same (and hence no i subscript was necessary).

Finally, note that it is possible to have multiple summations, which can be conducted in any order, so for example

$$\sum_{i=1}^{n} \sum_{j=1}^{m} x_{ij}$$

would mean sum over all of the i and j subscripts, but we could either sum over the j's first for each value of i or sum over the i's first for each value of j. Usually, the convention is that the inner sum (in this case the one that runs over j from 1 to m would be conducted first – i.e., separately for each value of i).

1.5.9 Pi Notation

Similar to the use of sigma to denote sums, the pi operator (Π) is used to denote repeated multiplications. For example

$$\prod_{i=1}^{n} x_i = x_1 x_2 \ldots x_n \tag{1.15}$$

means 'multiply together all of the x_i for each value of i between the lower and upper limits'. It also follows that

$$\prod_{i=1}^{n} (cx_i) = c^n \prod_{i=1}^{n} x_i$$

For example, the product

$$\prod_{i=1}^{4} i^2$$

is equal to $1^2 \times 2^2 \times 3^2 \times 4^2 = 1 \times 4 \times 9 \times 16 = 576$.

Sometimes we need to calculate the *geometric mean* of a series. If the series contains n elements, this would mean taking the n^{th} root. For example, as we will see in Chapter 2, we would calculate the holding period return on an investment paying a return in each period (assume this is a year) i of r_i where there a total of n years as

$$\prod_{i=1}^{n} (1 + r_i) = (1 + r_1)(1 + r_2) \ldots (1 + r_n)$$

To calculate the average return in each year, we would take the geometric mean (i.e., the nth root) of this, as

$$\sqrt[n]{\prod_{i=1}^{n} (1 + r_i)}$$

and then we would subtract one at the end. A detailed illustration will be given in Section 2.6 of Chapter 2.

1.5.10 Functions of More than one Variable

All the examples we have examined so far in this section involve situations where y is a function of a single variable x, but it is also possible for y to be a function of several variables. Returning to the example in Table 1.1 to illustrate, we might suppose that grades (y) depend on hours of study (x_1) and hours of tutoring (x_2), so we would write

$$y = a + b_1 x_1 + b_2 x_2 \tag{1.16}$$

where a is still interpreted as an intercept, but there are now two slopes: b_1 measures how much y varies with changes in x_1 while b_2 measures how much y varies with changes in x_2. In order to plot such a function, we would need a three-dimensional representation. This notation will be very useful in later chapters when we examine relationships between many variables and we can continue to extend the model in exactly the same way according to how many variables we have included.

1.6 Differential Calculus

The effect of the *rate of change of one variable on the rate of change of another* is measured by a mathematical derivative. If the relationship between the two variables can be represented by a curve, the gradient of the curve will be this rate of change. Consider a variable y that is some function f of another variable x, i.e., $y = f(x)$. The derivative of y with respect to x is written

$$\frac{dy}{dx} = \frac{df(x)}{dx}$$

or sometimes written as

$$\frac{dy}{dx} = f'(x)$$

This term measures the instantaneous rate of change of y with respect to x, or in other words, the impact of an infinitesimally small change in x. Notice the difference between the notations Δy and dy – the former refers to a change in y of any size, whereas the latter refers specifically to an infinitesimally small change.

1.6.1 Differentiation: the Fundamentals

The basic rules of differentiation are as follows

1. The derivative of a constant is zero

$$\text{e.g., if } y = 10, \frac{dy}{dx} = 0$$

This is because $y = 10$ would be represented as a horizontal straight line on a graph of y against x, and therefore the gradient of this function is zero.

2. The derivative of a linear function is simply its slope

$$\text{e.g., if } y = 3x + 2, \frac{dy}{dx} = 3$$

But non-linear functions will have different gradients at each point along the curve. In effect, the gradient at each point is equal to the gradient of the

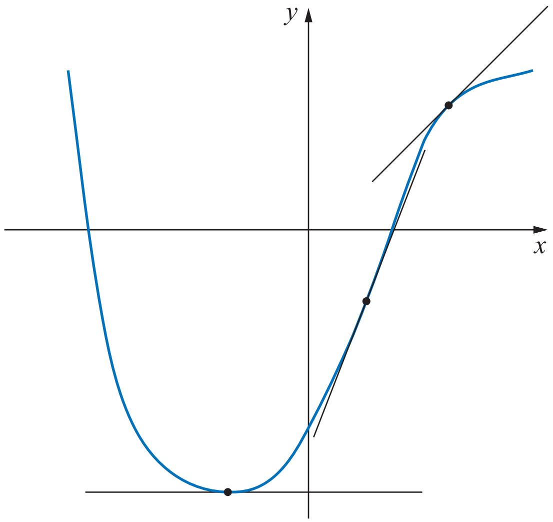


Figure 1.8 The tangents to a curve

tangent at that point – see Figure 1.8. Notice that the gradient will be zero at the point where the curve changes direction from positive to negative or from negative to positive – this is known as a *turning point* or equivalently as a *stationary point*.

3. The derivative of a power function n of x

$$\text{i.e., the derivative of } y = cx^n \text{ is given by } \frac{dy}{dx} = cnx^{n-1}$$

For example

$$y = 4x^3, \frac{dy}{dx} = (4 \times 3)x^2 = 12x^2$$

$$y = \frac{3}{x} = 3x^{-1}, \frac{dy}{dx} = (3 \times -1)x^{-2} = -3x^{-2} = \frac{-3}{x^2}$$

4. The derivative of a power of an entire function such as $[f(x)]^n$ is given by

$$\frac{dy}{dx} = n[f(x)]^{n-1}f'(x)$$

$$\text{e.g., if } y = (6x + x^4)^3, \frac{dy}{dx} = 3(6x + x^4)^2(6 + 4x^3)$$

5. The derivative of a sum is equal to the sum of the derivatives of the individual parts. Similarly, the derivative of a difference is equal to the difference of the derivatives of the individual parts

$$\text{e.g., if } y = f(x) + g(x), \frac{dy}{dx} = f'(x) + g'(x)$$

while

$$\text{if } y = f(x) - g(x), \frac{dy}{dx} = f'(x) - g'(x)$$

6. The derivative of the log of x is given by $1/x$

$$\text{i.e., } \frac{d(\ln(x))}{dx} = \frac{1}{x}$$

7. The derivative of the log of a function of x is the derivative of the function divided by the function

$$\text{i.e., } \frac{d(\ln(f(x)))}{dx} = \frac{f'(x)}{f(x)}$$

For example, the derivative of $\ln(x^3 + 2x - 1)$ is given by

$$\frac{d(\ln(x^3 + 2x - 1)}{dx} = \frac{3x^2 + 2}{x^3 + 2x - 1}$$

8. The derivative of an exponential of x is itself, so if $y = e^x$

$$\frac{dy}{dx} = e^x$$

More generally, the derivative of a function of an exponential is the derivative of the function multiplied by the exponential of the function, so if $y = e^{f(x)}$

$$\frac{dy}{dx} = f'(x)e^{f(x)}$$

So, to illustrate, if $y = e^{3x^2}$

$$\frac{dy}{dx} = 6xe^{3x^2}$$

1.6.2 Derivatives of Products and Quotients

Suppose that we have two functions multiplied together or one function divided by another function (recall that these are known as a *product* and a *quotient*, respectively). How would we differentiate these? Fortunately, both are fairly straightforward.

For a product, which could be written as $y = f(x)g(x)$, the rule is

$$\frac{dy}{dx} = f'(x)g(x) + f(x)g'(x)$$

For a quotient, written as $y = \frac{f(x)}{g(x)}$, the rule is

$$\frac{dy}{dx} = \frac{f'(x)g(x) - g'(x)f(x)}{g(x)^2}$$

Let us look at a couple of simple examples. Suppose that we have a product of two functions, $y = (3x^3 + 7x^2)(-2x^2 - 6)$. To differentiate this, product, we can view it as two functions, $y = f(x)g(x)$ and then we simply differentiate the first part, $f(x)$, multiplying that derivative by the second part, $g(x)$, unaltered, and then add the derivative of the second part multiplied by the first part unaltered

$$\frac{dy}{dx} = (9x^2 + 14x)(-2x^2 - 6) + (3x^3 + 7x^2)(-4x)$$

Again, it would be possible to simplify this expression but this is left as an exercise.

Now, suppose the quotient that we wish to differentiate is

$$y = \frac{(6x^4 - x)}{(3x^3 - 2x^2 + 4)}$$

Following the quotient rule, the derivative would be

$$\frac{dy}{dx} = \frac{(24x^3 - 1)(3x^3 - 2x^2 + 4) + (9x^2 - 4x)(6x^4 - x)}{(3x^3 - 2x^2 + 4)^2}$$

1.6.3 Higher Order Derivatives

It is possible to differentiate a function more than once to calculate the second order, third order, ..., nth order derivatives. The notation for the second order derivative (which is usually just termed the second derivative, and which is the highest order derivative that we will need in this book) is

$$\frac{d^2y}{dx^2} = f''(x) = \frac{d(\frac{dy}{dx})}{dx}$$

To calculate second order derivatives, we simply differentiate the function with respect to x and then we differentiate it again. For example, suppose that we have the function

$$y = 4x^5 + 3x^3 + 2x + 6$$

The first order derivative is

$$\frac{dy}{dx} = \frac{d(4x^5 + 3x^3 + 2x + 6)}{dx} = f'(x) = 20x^4 + 9x^2 + 2$$

The second order derivative is

$$\frac{d^2y}{dx^2} = f''(x) = \frac{d(\frac{d(4x^5+3x^3+2x+6)}{dx})}{dx} = \frac{d(20x^4 + 9x^2 + 2)}{dx} = 80x^3 + 18x$$

The second order derivative can be interpreted as the gradient of the gradient of a function – i.e., the rate of change of the gradient.

We said above that at the turning point of a function its gradient will be zero. How can we tell, then, whether a particular turning point is a maximum or a minimum? In other words, is the shape of the function for that value of x a ∪ or a ∩? The answer is that to do this we would look at the second derivative. When a function reaches a maximum, its second derivative is negative, while it is positive for a minimum.

For example, consider the quadratic function $y = 5x^2 + 3x - 6$. We already know that since the squared term in the equation has a positive sign (i.e., it is 5 rather than, say, -5), the function will have a ∪-shape rather than an ∩-shape, and thus it will have a minimum rather than a maximum. But let us also demonstrate this using differentiation

$$\frac{dy}{dx} = 10x + 3, \frac{d^2y}{dx^2} = 10$$

Since the second derivative is positive, the function indeed has a minimum because the rate of change of the slope is positive as the gradient switches from negative on the left of the minimum to zero at the minimum to positive on the right of the minimum.

To find where this minimum is located, take the first derivative, set it to zero and solve it for x. So we have $10x + 3 = 0$, and thus $x = -\frac{3}{10} = -0.3$. If $x = -0.3$, the corresponding value of y is found by substituting -0.3 into the original function $y = 5x^2 + 3x - 6 = 5 \times (-0.3)^2 + (3 \times -0.3) - 6 = -6.81$. Therefore, the minimum of this function is found at $(-0.3, -6.81)$.

What if the second derivative of a function is zero for a particular value of x? In such cases, the function is at a *point of inflection*. Turning points and points of inflection are both types of stationary point. At a point of inflection, the figure has neither a ∪-shape or a ∩-shape but something more like an 'S'.

To illustrate, consider the function $y = (x + 6)^3 - 5$. Its first derivative is $f'(x) = 3(x + 6)^2$. The second derivative is $f''(x) = 6(x + 6)$. Suppose that we are interested in evaluating the shape of the function at $x = -6$. At this point, $y = -5$, $f'(x) = 0$ and $f'' = 0$ so this is a point of inflection. We plot the original function, $y = f(x)$, the first derivative function, $y = f'(x)$, and the second derivative function, $y = f''(x)$, in Figure 1.9.

1.6.4 Differentiation of Functions of Functions Using the Chain Rule

In the section above we saw how to differentiate powers of functions and logarithms of functions. These are just special cases of a more general situation where we might want to differentiate a function of a function, $y = f(g(x))$. In such situations, we effectively split the process into two parts: we differentiate y with respect to g and then multiply it by the derivative of g with respect to x. We can write this as

$$\frac{dy}{dx} = \frac{dy}{dg}\frac{dg}{dx}$$

It is easy to see why this approach is often known as the *chain rule of differentiation*. As an illustration, suppose that we wish to differentiate the function $y = (4x^3 - 6x + 4)^4$. In this case we would have $g(x) = 4x^3 - 6x + 4$ and $y = g^4$. The derivative of y

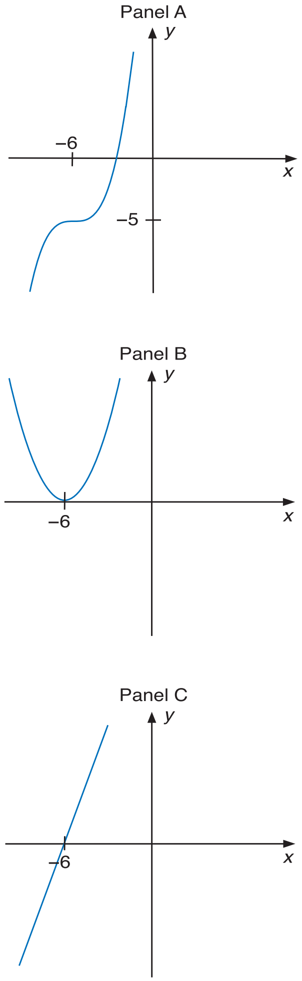


Figure 1.9 $y = f(x)$, its first derivative and its second derivative around the point $x = -6$

with respect to g is

$$\frac{dy}{dg} = 4g^3$$

and the derivative of g with respect to x is

$$\frac{dg}{dx} = 12x^2 - 6$$

Putting these together, the derivative of y with respect to x is:

$$\frac{dy}{dx} = \frac{dy}{dg}\frac{dg}{dx} = (4g^3)(12x^2 - 6) = 4(4x^3 - 6x + 4)^3(12x^2 - 6)$$

It may be possible to simplify this function but we leave it in its factorised form.

EXAMPLE 1.2: Utility Functions

In economics, utility provides a measure of the satisfaction that a consumer derives from a good or service that they have purchased. In finance, the concept is usually used to measure how satisfaction changes with differing levels of (terminal, i.e., end of period) wealth or with risk and return. Utilities constitute a useful illustration of how the concepts of functions and differentiation can be applied in finance.

Let us start by considering utility as a function of wealth. We would write the utility function as $U = f(W)$. Many such utility functions would be possible, e.g.

1. $U = 5 + 8W$
2. $U = 30 - e^{0.5W}$
3. $U = 100W + 0.5W^2$
4. $U = \ln(W)$

But would they all make sense as utility functions and what are the properties that we would want a utility function to have?

SOLUTION For a utility function to be plausible, we usually have two requirements. First, we would want to ensure that the investor has a positive marginal utility of wealth – in other words, utility always rises with wealth or mathematically, $dU/dW > 0$. We would also usually expect that investors are *risk averse* – i.e., they would reject a fair gamble or they prefer less risk to more. When utility is a function of wealth, it turns out that the condition for the investor to be risk averse is $d^2U/dW^2 < 0$. This condition also implies that marginal utility diminishes with wealth – in other words, I get more utility the more wealth I have, but each additional unit of wealth gives me less and less additional satisfaction. This makes intuitive sense.

For completeness, note that a *risk neutral* investor would be indifferent to a gamble and the second derivative of their utility function with respect to wealth would be: $d^2U/dW^2 = 0$. A *risk loving* investor who would prefer more risk to less and who would therefore accept a fair gamble has a second derivative greater than zero: $d^2U/dW^2 > 0$.

So to evaluate the plausibility of each of the four utility functions above, we would need to differentiate each of them twice and determine whether the first derivative is positive and the second derivative negative. These would be:

1. $U = 5 + 8W, \quad dU/DW = 8, \quad d^2U/DW^2 = 0$
2. $U = 30 - 30e^{0.5W}, \quad dU/DW = -15e^{0.5W}, \quad d^2U/DW^2 = -7.5e^{0.5W}$
3. $U = 100W + 0.5W^2, \quad dU/DW = 100 + W, \quad d^2U/DW^2 = 1$
4. $U = \ln(W), \quad dU/DW = 1/W, \quad d^2U/DW^2 = -1/W^2$

Utility function 1 is a linear equation, sloping upwards. The first derivative is positive for all values of W and so the investor having this utility function would have a positive marginal utility of wealth but the second derivative is zero and thus such an investor would be risk neutral.

Utility function 2 has a first derivative that is negative (so that the investor prefers less wealth to more) and a second derivative that is also negative for all values of W since e^{ax} will be positive for any (positive or negative) value of a and thus the investor would be risk averse.

The third utility function has a first derivative that is positive for any value of W greater than -100, and a second derivative that is positive everywhere and thus the investor would be risk loving.

Finally, the fourth utility function has a first derivative that is positive for all positive values of W but a negative second derivative for all values of W. So overall, we would conclude that the fourth utility function is the most appropriate of the four to describe a typical investor as it is the only one having the required properties of a positive first derivative and a negative second derivative.

1.6.5 Partial Differentiation

In the case where y is a function of more than one variable (e.g., $y = f(x_1, x_2, ..., x_n)$), it may be of interest to determine the effect that changes in each of the individual x variables would have on y. The differentiation of y with respect to only one of the variables, holding the others constant, is known as *partial differentiation*. The partial derivative of y with respect to a variable x_1 is usually denoted

$$\frac{\partial y}{\partial x_1}$$

All of the rules for differentiation explained above still apply and there will be one (first order) partial derivative for each variable on the RHS of the equation. We calculate these partial derivatives one at a time, treating all of the other variables as if they were constants. To give an illustration, suppose $y = 3x_1^3 + 4x_1 - 2x_2^4 + 2x_2^2$. The partial derivative of y with respect to x_1 would be

$$\frac{\partial y}{\partial x_1} = 9x_1^2 + 4$$

while the partial derivative of y with respect to x_2 would be

$$\frac{\partial y}{\partial x_2} = -8x_2^3 + 4x_2$$

As we will see in Chapter 3, the ordinary least squares (OLS) estimator gives formulae for the values of the parameters that minimise the function given by $L = \sum_t (y_t - \hat{\alpha} - \hat{\beta} x_t)^2$. The minimum of L (the residual sum of squares) is found by partially differentiating this function with respect to $\hat{\alpha}$ and then separately with respect to $\hat{\beta}$ and setting these partial derivatives to zero. Therefore, partial differentiation has a key role in deriving the main approach to parameter estimation that we use in econometrics – see Appendix 3.1 at the end of Chapter 3 for a demonstration of this application.

1.6.6 Functions that Cannot be Differentiated

Fortunately, it is possible to differentiate the majority of functions of interest to us in finance, but are there any formulations where it is not possible to calculate the gradient? The answer is that there are particular difficulties where a function is *discontinuous* or, in other words, it contains a jump (either up or down). For example, if we have a function $y = f(x)$ which takes a certain form when y is positive or zero and a different form when y is negative such as

$$y = \begin{cases} 2x + 4 & if\, x \geq 0 \\ -x + 3 & if\, x < 0 \end{cases} \tag{1.17}$$

It would not be possible to differentiate this function, which is known as a piecewise linear model, since each of the pieces (≥ 0 and < 0) are linear functions of x but overall it is non-linear. These models will be discussed in more detail in Chapter 10.

1.6.7 Derivatives in Use in Finance

What do we actually use differentiation for? A key use relates to the concept of what happens *at the margin* – in other words, what is the effect of an infinitesimally small change in x on y – this is exactly the interpretation of the slope of a function at a specific value of x. In reality, we usually weaken this slightly to say that the derivative of y with respect to x can be used to measure the effect of a unit change in x on y. This is a very useful concept that is widely used in measuring marginal utility, marginal propensity to save as income changes, etc. – for instance, what is the effect of a one-unit change in wealth upon the utility of an investor?

Differentiation relates *unit* changes in x to *unit* changes in y but it will often be of interest to consider what happens to y if x changes by one *percent* rather than one unit. This would be measured by an *elasticity*. The formula for calculating an elasticity of y with respect to x would be given by

$$\text{elasticity} = \frac{dy}{dx}\frac{x}{y} \tag{1.18}$$

EXAMPLE 1.3

Suppose that the demand for an on-line stock brokerage account is given by the following function

$$q = 100{,}000 - 500p$$

where q is the number of trades made per month and p is the fee charged per trade. If $p =$£20, calculate the price elasticity of demand.

SOLUTION To solve this, we first need to calculate the derivative of q with respect to p, which is very straightforward as it is a linear function: $dq/dp = -500$. To then calculate the elasticity, we need to calculate the value of q that corresponds to the value of p in the question (20). If $p = 20$, $q = 100{,}000 - (500 \times 20) = 90{,}000$. We then calculate the elasticity as

$$\text{elasticity} = \frac{dq}{dp}\frac{p}{q} = -500 \times \frac{20}{90{,}000} = -0.111$$

This would be interpreted as implying that a 1% increase in the fee per trade would reduce the number of trades by 0.111%. Since this figure is less than one in absolute value, we would conclude that demand for brokerage services is *inelastic* and thus the firm may have the opportunity to increase its revenue and profits by raising prices.

1.6.8 Integration

Integration is the opposite of differentiation, so that if we integrate a function and then differentiate the result, we get back the original function. Recall that derivatives give functions for calculating the slope of a curve; integration, on the other hand, is used to calculate the area under a curve (between two specific points). Further details on the rules for integration are beyond the scope of this book since the mathematical technique is not needed for any of the econometric approaches we will employ, but it will be useful to be familiar with the general concept. For further reading, see for example Renshaw (2016, Chapter 18).

1.7 Matrices

Before we can work with matrices, we need to define some terminology and to distinguish between a scalar, a vector and a matrix.

- A *scalar* is simply a single number (although it need not be a whole number – e.g., 3, −5, 0.5 are all scalars)
- A *vector* is a one-dimensional *array of numbers* (see below for examples)
- A *matrix* is a two-dimensional *collection or array of numbers*. The size of a matrix is given by its numbers of rows and columns.

Matrices are very useful and important ways for organising sets of data together, which make manipulating and transforming them much easier than it would be to work with each constituent of the matrix separately. Matrices are widely used in econometrics and finance for solving systems of linear equations, for deriving key results and for expressing formulae in a succinct way. Sometimes **bold-faced type** is used to denote a vector or matrix (e.g., A), although in this book we will not do so – hopefully it should be obvious whether an object is a scalar, vector or matrix from the context, or this will be clearly stated. Some useful features of matrices and explanations of how to work with them are now described.

- The dimensions of a matrix are quoted as $R \times C$, which is the number of rows by the number of columns.
- Each element in a matrix is referred to using subscripts. For example, suppose a matrix M has two rows and four columns. The element in the second row and the third column of this matrix would be denoted m_{23}, so that more generally m_{ij} refers to the element in the ith row and the jth column. Thus a 2×4 matrix would have elements

$$\begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \end{pmatrix}$$

- Vectors are special cases of matrices where there is only one column or only one row.
- If a matrix has only one row, it is known as a *row vector*, which will be of dimension $1 \times C$, where C is the number of columns

$$\text{e.g.,} \quad (2.7 \quad 3.0 \quad -1.5 \quad 0.3)$$

- A matrix having only one column is known as a *column vector*, which will be of dimension $R \times 1$, where R is the number of rows

$$\text{e.g.,} \quad \begin{pmatrix} 1.3 \\ -0.1 \\ 0.0 \end{pmatrix}$$

- When the number of rows and columns is equal (i.e., $R = C$), it would be said that the matrix is square, as is the following 2×2 matrix

$$\begin{pmatrix} 0.3 & 0.6 \\ -0.1 & 0.7 \end{pmatrix}$$

- A matrix in which all the elements are zero is known as a *zero matrix*

$$\text{e.g.,} \quad \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

- A *symmetric matrix* is a special type of square matrix that is symmetric about the leading diagonal (the diagonal line running through the matrix from the top left to the bottom right), so that $m_{ij} = m_{ji} \;\forall\; i, j$

$$\text{e.g.,} \quad \begin{pmatrix} 1 & 2 & 4 & 7 \\ 2 & -3 & 6 & 9 \\ 4 & 6 & 2 & -8 \\ 7 & 9 & -8 & 0 \end{pmatrix}$$

- A diagonal matrix is a square matrix which has non-zero terms on the leading diagonal and zeros everywhere else

$$\text{e.g.,}\quad \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

- A diagonal matrix with 1 in all places on the leading diagonal and zero everywhere else is known as the *identity matrix*, denoted by I. By definition, an identity matrix must be symmetric (and therefore also square)

$$\text{e.g.,}\quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- The identity matrix is essentially the matrix equivalent of the number one. Multiplying any matrix by the identity matrix of the appropriate size results in the original matrix being left unchanged. So for any matrix M

$$MI = IM = M$$

1.7.1 Operations with Matrices

In order to perform operations with matrices (e.g., addition, subtraction or multiplication), the matrices concerned must be *conformable*. The dimensions of matrices required for them to be conformable depend on the operation.

- Addition and subtraction of matrices requires the matrices concerned to be of the same order (i.e., to have the same number of rows and the same number of columns as one another). The operations are then performed element by element

$$\text{e.g.,}\quad \text{if } A = \begin{pmatrix} 0.3 & 0.6 \\ -0.1 & 0.7 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0.2 & -0.1 \\ 0 & 0.3 \end{pmatrix}$$

$$A + B = \begin{pmatrix} 0.3 + 0.2 & 0.6 - 0.1 \\ -0.1 + 0 & 0.7 + 0.3 \end{pmatrix} = \begin{pmatrix} 0.5 & 0.5 \\ -0.1 & 1.0 \end{pmatrix}$$

$$A - B = \begin{pmatrix} 0.3 - 0.2 & 0.6 - -0.1 \\ -0.1 - 0 & 0.7 - 0.3 \end{pmatrix} = \begin{pmatrix} 0.1 & 0.7 \\ -0.1 & 0.4 \end{pmatrix}$$

- Multiplying or dividing a matrix by a scalar (that is, a single number), implies that every element of the matrix is multiplied by that number

$$\text{e.g.,}\quad 2A = 2\begin{pmatrix} 0.3 & 0.6 \\ -0.1 & 0.7 \end{pmatrix} = \begin{pmatrix} 0.6 & 1.2 \\ -0.2 & 1.4 \end{pmatrix}$$

- More generally, for two matrices A and B of the same order and for c a scalar, the following results hold

$$A + B = B + A$$

$$A + 0 = 0 + A = A$$

$$cA = Ac$$

$$c(A + B) = cA + cB$$

$$A0 = 0A = 0$$

- Multiplying two matrices together requires the number of columns of the first matrix to be equal to the number of rows of the second matrix. Note also that the ordering of the matrices is important when multiplying them, so that in general, $AB \neq BA$. When matrices are multiplied together, the resulting matrix will be of size (number of rows of first matrix × number of columns of second matrix), e.g., if we multiply a (3×2) matrix by a (2×4) matrix, the result is a (3×4) matrix: $(3 \times 2) \times (2 \times 4) = (3 \times 4)$. In terms of determining the dimensions of the matrix, it is as if the number of columns of the first matrix and the number of rows of the second cancel out.[1] This rule also follows more generally, so that $(a \times b) \times (b \times c) \times (c \times d) \times (d \times e) = (a \times e)$, etc.
- The actual multiplication of the elements of the two matrices is done by multiplying along the rows of the first matrix and down the columns of the second

$$\text{e.g.,}\quad \begin{pmatrix} 1 & 2 \\ 7 & 3 \\ 1 & 6 \end{pmatrix} \begin{pmatrix} 0 & 2 & 4 & 9 \\ 6 & 3 & 0 & 2 \end{pmatrix}$$

$$(3 \times 2) \quad (2 \times 4)$$

$$= \begin{pmatrix} ((1 \times 0) + (2 \times 6)) & ((1 \times 2) + (2 \times 3)) & ((1 \times 4) + (2 \times 0)) & ((1 \times 9) + (2 \times 2)) \\ ((7 \times 0) + (3 \times 6)) & ((7 \times 2) + (3 \times 3)) & ((7 \times 4) + (3 \times 0)) & ((7 \times 9) + (3 \times 2)) \\ ((1 \times 0) + (6 \times 6)) & ((1 \times 2) + (6 \times 3)) & ((1 \times 4) + (6 \times 0)) & ((1 \times 9) + (6 \times 2)) \end{pmatrix}$$

$$(3 \times 4)$$

$$= \begin{pmatrix} 12 & 8 & 4 & 13 \\ 18 & 23 & 28 & 69 \\ 36 & 20 & 4 & 21 \end{pmatrix}$$

$$(3 \times 4)$$

---

[1] Of course, the actual elements of the matrices themselves do not cancel out – this is just a simple rule of thumb for calculating the dimensions of the matrix resulting from a multiplication.

- In general, matrices cannot be divided by one another. Instead, we achieve the same sort of outcome by multiplying by the inverse – see below.
- The transpose of a matrix, written A' or A^{T}, is the matrix obtained by transposing (switching) the rows and columns of a matrix

$$\text{e.g.,} \quad \text{if } A = \begin{pmatrix} 1 & 2 \\ 7 & 3 \\ 1 & 6 \end{pmatrix} \quad \text{then } A' = \begin{pmatrix} 1 & 7 & 1 \\ 2 & 3 & 6 \end{pmatrix}$$

If A is of dimensions $R \times C$, A' will be $C \times R$.

1.7.2 The Rank of a Matrix

The rank of a matrix A is given by the maximum number of linearly independent rows (or columns) contained in the matrix. For example,

$$\operatorname{rank}\begin{pmatrix} 3 & 4 \\ 7 & 9 \end{pmatrix} = 2$$

since both rows and columns are (linearly) independent of one another, but

$$\operatorname{rank}\begin{pmatrix} 3 & 6 \\ 2 & 4 \end{pmatrix} = 1$$

as the second column is not independent of the first (the second column is simply twice the first and also the second row is two thirds of the first). A matrix with a rank equal to its dimension, as in the first of these two cases, is known as a *matrix of full rank*. A matrix that is less than of full rank is known as a *short rank matrix*, and such a matrix is also termed *singular*.

Three important results concerning the rank of a matrix are

- $\text{Rank}(A) = \text{Rank}(A')$
- $\text{Rank}(AB) \leq \min(\text{Rank}(A), \text{Rank}(B))$
- $\text{Rank}(A'A) = \text{Rank}(AA') = \text{Rank}(A)$

1.7.3 The Inverse of a Matrix

The inverse of a matrix A, where defined, is denoted A^{-1}. It is that matrix which, when pre-multiplied or post-multiplied by A, will result in the identity matrix

$$\text{i.e.,} \quad AA^{-1} = A^{-1}A = I$$

The inverse of a matrix exists only when the matrix is square and non-singular – that is, when it is of full rank. The inverse of a 2×2 non-singular matrix whose elements are

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

will be given by

$$\frac{1}{ad - bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

The expression in the denominator above to the left of the matrix $(ad - bc)$ is the *determinant* of the matrix, and will be a scalar. If this determinant is zero, the matrix is *singular*, and thus not of *full rank* so that its inverse does not exist. For example, if

$$A = \begin{pmatrix} 1 & 6 \\ 2 & 12 \end{pmatrix}$$

$ad - bc = 12 - 12 = 0$ so this matrix is singular since the second column is six times the first (or looking at it another way, the second row is double the first). We usually write the determinant of a matrix using $|.|$ (the same notation as for the absolute value of a variable). So $|A|$ is the determinant of matrix A.

EXAMPLE 1.4

If the matrix is

$$\begin{pmatrix} 2 & 1 \\ 4 & 6 \end{pmatrix}$$

the inverse will be

$$\frac{1}{8}\begin{pmatrix} 6 & -1 \\ -4 & 2 \end{pmatrix} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{8} \\ -\frac{1}{2} & \frac{1}{4} \end{pmatrix}$$

As a check, multiply the two matrices together and it should give the identity matrix – the matrix equivalent of one (analogous to $\frac{1}{3} \times 3 = 1$)

$$\begin{pmatrix} 2 & 1 \\ 4 & 6 \end{pmatrix} \times \frac{1}{8}\begin{pmatrix} 6 & -1 \\ -4 & 2 \end{pmatrix} = \frac{1}{8}\begin{pmatrix} 8 & 0 \\ 0 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$= I$, as required.

The calculation of the inverse of an $N \times N$ matrix for $N > 2$ is more complex. Two of the most common approaches to finding the inverse of a larger matrix are known as the *method of determinants* and the *Gauss-Jordan elimination method*. These are beyond the scope of this text but see Wisniewski (2013), for example, for further details.

Properties of the inverse of a matrix include

- $I^{-1} = I$
- $(A^{-1})^{-1} = A$
- $(A')^{-1} = (A^{-1})'$
- $(AB)^{-1} = B^{-1}A^{-1}$

1.7.4 The Trace of a Matrix

The trace of a square matrix is the sum of the terms on its leading diagonal. For example, the trace of the matrix

$$A = \begin{pmatrix} 3 & 4 \\ 7 & 9 \end{pmatrix}$$

written Tr(A), is $3 + 9 = 12$. Some important properties of the trace of a matrix are

- $\mathrm{Tr}(cA) = c\mathrm{Tr}(A)$
- $\mathrm{Tr}(A') = \mathrm{Tr}(A)$
- $\mathrm{Tr}(A + B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$
- $\mathrm{Tr}(I_N) = N$

1.7.5 The Eigenvalues of a Matrix

The concept of the eigenvalues of a matrix is necessary for testing for long-run relationships between series using what is known as the Johansen cointegration test used in Chapter 8. Let Π denote a $p \times p$ square matrix, c denote a $p \times 1$ non-zero vector, and λ denote a set of scalars. λ is called a *characteristic root* or set of roots of the matrix Π if it is possible to write

$$\underset{p\times p}{\Pi}\,\underset{p\times 1}{c} = \underset{p\times 1}{\lambda c}$$

This equation can also be written as

$$\Pi c = \lambda I_p c$$

where I_p is an identity matrix, and hence

$$(\Pi - \lambda I_p)c = 0$$

Since $c \neq 0$ by definition, then for this system to have a non-zero solution, the matrix $(\Pi - \lambda I_p)$ is required to be singular (i.e., to have a zero determinant)

$$|\Pi - \lambda I_p| = 0$$

For example, let Π be the 2×2 matrix

$$\Pi = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix}$$

Then the characteristic equation is

$$\begin{aligned}
&|\Pi - \lambda I_p| \\
&\quad = \left| \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0 \\
&\quad = \begin{vmatrix} 5-\lambda & 1 \\ 2 & 4-\lambda \end{vmatrix} = (5-\lambda)(4-\lambda) - 2 = \lambda^2 - 9\lambda + 18
\end{aligned}$$

This gives the solutions $\lambda = 6$ and $\lambda = 3$. The characteristic roots are also known as *eigenvalues*. The eigenvectors would be the values of c corresponding to the eigenvalues. Some properties of the eigenvalues of any square matrix A are

- the sum of the eigenvalues is the trace of the matrix
- the product of the eigenvalues is the determinant
- the number of non-zero eigenvalues is the rank

For a further illustration of the last of these properties, consider the matrix

$$\Pi = \begin{bmatrix} 0.5 & 0.25 \\ 0.7 & 0.35 \end{bmatrix}$$

Its characteristic equation is

$$\left| \begin{bmatrix} 0.5 & 0.25 \\ 0.7 & 0.35 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right| = 0$$

which implies that

$$\begin{vmatrix} 0.5-\lambda & 0.25 \\ 0.7 & 0.35-\lambda \end{vmatrix} = 0$$

This determinant can also be written $(0.5-\lambda)(0.35-\lambda) - (0.7 \times 0.25) = 0$ or

$$0.175 - 0.85\lambda + \lambda^2 - 0.175 = 0$$

or

$$\lambda^2 - 0.85\lambda = 0$$

which can be factorised to $\lambda\,(\lambda - 0.85) = 0$.

The characteristic roots are therefore 0 and 0.85. Since one of these eigenvalues is zero, it is obvious that the matrix Π cannot be of full rank. In fact, this is also obvious from just looking at Π, since the second column is exactly half the first.

KEY CONCEPTS

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

- functions
- turning points
- the chain rule
- powers
- exponentials
- sigma and pi notation
- quadratic equations
- inverse of a matrix
- eigenvalues
- roots
- derivatives and differentiation
- products and quotients
- indices
- polynomials
- logarithms
- conformable matrix
- rank of a matrix
- eigenvectors

SELF-STUDY QUESTIONS

1. (a) If $f(x) = 3x^2 - 4x + 2$, find $f(0), f(2), f(-1)$
  (b) If $f(x) = 4x^2 + 2x - 3$, find $f(0), f(3), f(a), f(3 + a)$
  (c) Considering your answers to the previous question part, in general does $f(a) + f(b) = f(a + b)$? Explain.

2. Simplify the following as much as possible
  (a) $4x^5 \times 6x^3$
  (b) $3x^2 \times 4y^2 \times 8x^4 \times -2y^4$
  (c) $(4p^2q^3)^3$
  (d) $6x^5 \div 3x^2$
  (e) $7y^2 \div 2y^5$
  (f) $\dfrac{3(xy)^3 \times 6(xz)^4}{2(xy)^2x^3}$
  (g) $(xy)^3 \div x^3y^3$
  (h) $(xy)^3 - x^3y^3$

3. Solve the following
  (a) $125^{1/3}$
  (b) $64^{1/3}$
  (c) $16^{1/4}$
  (d) $9^{3/2}$
  (e) $9^{2/3}$
  (f) $81^{1/2} + 64^{1/2} + 64^{1/3}$

4. Write each of the following as a prime number raised to a power
  (a) 9
  (b) 625
  (c) 125^{-1}

5. Solve the following equations
  (a) $3x - 6 = 6x - 12$
  (b) $2x - 304x + 8 = x + 9 - 3x + 4$
  (c) $\frac{x+3}{2} = \frac{2x-6}{3}$

6. Write out all of the terms in the following and evaluate them
  (a) $\sum_{j=1}^{3} j$
  (b) $\sum_{j=2}^{5} (j^2 + j + 3)$
  (c) $\sum_{i=1}^{n} x$ with $n = 4$ and $x = 3$
  (d) $\prod_{j=1}^{3} x$ with $x = 2$
  (e) $\prod_{i=3}^{6} i$

7. Write the equations for each of the following lines
  (a) Gradient $=$ 3, intercept $= -1$
  (b) Gradient $= -2$, intercept $=$ 4
  (c) Gradient $= \frac{1}{2}$, crosses y-axis at 3

(d) Gradient $= \frac{1}{2}$, crosses x-axis at 3
(e) Intercept 2 and passing through (3,1)
(f) Gradient 4 and passing through $(-2,-2)$
(g) Passes through $x = 4, y = 2$ and $x = -2, y = 6$

8. Differentiate the following functions twice with respect to x
(a) $y = 6x$
(b) $y = 3x^2 + 2$
(c) $y = 4x^3 + 10$
(d) $y = \frac{1}{x}$
(e) $y = x$
(f) $y = 7$
(g) $y = 6x^{-3} + \frac{6}{x^3}$
(h) $y = 3 \ln x$
(i) $y = \ln(3x^2)$
(j) $y = \dfrac{3x^4 - 6x^2 - x - 4}{x^3}$

9. Differentiate the following functions partially with respect to x and (separately) partially with respect to y
(a) $z = 10x^3 + 6y^2 - 7y$
(b) $z = 10xy^2 - 6$
(c) $z = 6x$
(d) $z = 4$

10. Factorise the following expressions
(a) $x^2 - 7x - 8$
(b) $5x - 2x^2$
(c) $2x^2 - x - 3$
(d) $6 + 5x - 4x^2$
(e) $54 - 15x - 25x^2$

11. Express the following in logarithmic form
(a) $5^3 = 125$
(b) $11^2 = 121$
(c) $6^4 = 1296$

12. Evaluate the following (without using a calculator)
(a) $\ln_{10} 10{,}000$
(b) $\ln_2 16$
(c) $\ln_{10} 0.01$
(d) $\ln_5 125$
(e) $\ln_e e^2$

13. Express the following logarithms using powers
(a) $\ln_5 3125 = 5$
(b) $\ln_{49} 7 = \frac{1}{2}$
(c) $\ln_{0.5} 8 = -3$

14. Write the following as simply as possible as sums of logs of prime numbers

(a) $\ln 60$

(b) $\ln 300$

15. Simplify the following as far as possible

(a) $\ln 27 - \ln 9 + \ln 81$

(b) $\ln 8 - \ln 4 + \ln 32$

16. Solve the following

(a) $\ln x^4 - \ln x^3 = \ln 5x - \ln 2x$

(b) $\ln(x-1) + \ln(x+1) = 2\ln(x+2)$

(c) $\log_{10} x = 4$

17. Use the result that ln(8) is approximately 2.1 to estimate the following (without using a calculator):

(a) $\ln(16)$

(b) $\ln(64)$

(c) $\ln(4)$

18. Solve the following using logs and a calculator

(a) $4^x = 6$

(b) $4^{2x} = 3$

(c) $3^{2x-1} = 8$

19. Find the minima of the following functions. In each case, state the value of the function at the minimum

(a) $y = 6x^2 - 10x - 8$

(b) $y = (6x^2 - 8)^2$

20. Construct an example not used elsewhere in this book to demonstrate that for two conformable matrices A and B, $(AB)^{-1} = B^{-1}A^{-1}$.

21. Suppose that we have the following four matrices

$$A = \begin{bmatrix} 1 & 6 \\ -2 & 4 \end{bmatrix}, B = \begin{bmatrix} -3 & -8 \\ 6 & 4 \end{bmatrix}, C = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, D = \begin{bmatrix} 6 & -2 \\ 0 & -1 \\ 3 & 0 \end{bmatrix}$$

(a) Which pairs of matrices can be validly multiplied together? For these pairs, perform the multiplications.

(b) Calculate $2A$, $3B$, $\frac{1}{2}D$

(c) Calculate $\mathrm{Tr}(A)$, $\mathrm{Tr}(B)$, $\mathrm{Tr}(A+B)$ and verify that $\mathrm{Tr}(A) + \mathrm{Tr}(B) = \mathrm{Tr}(A+B)$

(d) What is the rank of the matrix A?

(e) Find the eigenvalues of the matrix $(A+B)$

(f) What will be the trace of the identity matrix of order 12?

22. (a) Add

$$\begin{bmatrix} 2 & -1 \\ -7 & 4 \end{bmatrix} \text{ to } \begin{bmatrix} -3 & 0 \\ 7 & -4 \end{bmatrix}$$