Preface

Financial econometrics is an exciting young discipline that began to take on its present form around the turn of the millennium. The subject brings financial theory and econometric methods together with the power of data to advance our understanding of the global financial universe upon which all modern economies depend. Two major developments underscored its rapid growth and expanding capabilities. First, the massive importance of well-functioning financial markets to the global economy and to global financial stability was universally acknowledged following the dot-com bubble of the late 1990s in the United States and the global financial crisis of 2008 coupled with its prolonged aftermath. Second, modern methods of econometrics emerged that proved equal to some of the special challenges presented by financial data and the ideas of financial theory.

Among the most significant of these challenges are the complex interdependencies of financial, commodity, and real estate markets, the dynamic and spatial linkages within financial data, the random wandering behaviour of asset prices, anomalies such as financial bubbles and market crashes in the data, the difficulties in modelling rapidly changing volatility in financial returns, the growth in high dimensional ultra-high frequency data, and the attention to market microstructure effects that all such data require. While not entirely unique to financial data, these challenges presented the econometrics profession with the need to re-fashion methods, develop new tools of inference, and tackle a wide selection of new empirical goals associated with a growing number of financial instruments and vast data sets being generated in the financial world.

This book, like the subject itself, is motivated by all of these challenges. We seek to provide a broad and gentle introduction to this rapidly developing subject of financial econometrics where theory, measurement, and data play equal roles in our development and where empirical applications occupy a central position. Our target audiences are intermediate and advanced undergraduate students, honours students who wish to learn about financial econometrics, and postgraduate students with limited backgrounds in econometrics who are doing masters courses designed to offer an introduction to finance and its applications. We hope the book will also prove useful to practitioners in the industry as an introductory reference source for relevant tools and approaches to modern empirical work in finance.

Throughout the book special emphasis is placed on the exposition of core concepts, their illustration using relevant financial data sets, and a hands-on approach to learning by doing that involves practical implementation. The guiding principle we have adopted is that only by working through plenty of applications and exercises can a coherent understanding of the properties of financial econometric models, interrelationships with the underlying finance theory, and the role of econometric tools of inference be achieved.
Our philosophy has been to write a book on financial econometrics, not an econometrics text that illustrates techniques with datasets drawn from finance. Our goal is centred on the subject of financial econometrics explaining how evidenced-based research in applied finance is conducted. Econometrics is viewed as the vehicle that makes the ideas and theories about financial markets face the reality of observations.

To ensure the book is self contained for a first course in financial econometrics, some foundational theory and methods of relevant econometric technique are provided. But the methods covered in this book travel along a customised path designed to ease the reader’s transition from concepts and methods to empirical work. The book tracks its way forward from data to modelling through to estimation, inference, and prediction.

A consistent thematic throughout the book is to motivate each topic with the presentation of relevant data. The journey begins with data and a simple grounding in regression and inference. From this foundation, it moves on to more advanced financial econometric methods that open up empirical applications with many different types of data from various financial markets. The path promises to keep readers motivated throughout their journey by means of many examples and to reinforce their learning by extensive data-based exercises. Several introductory Appendices are included to assist students with limited mathematics and no econometric background in understanding more technical aspects of the discussion particularly in the second half of the book.

Organization of the Book

Part I – Fundamentals – is designed to form the basis of a semester long first course in financial econometrics directed at an introductory level. Technical difficulty is kept to an absolute minimum with an emphasis on the data, financial concepts, appropriate econometric methodology, and the intuition that draws these essential components of modelling together. Methodology is largely confined to descriptive methods and ordinary least squares regression, a choice that limits the extent of the analysis and promotes heuristic discussion on some topics which are revisited later in the book for a more complete and rigorous development.

In Part II – Methods – the level of difficulty steps up slightly in treating the relevant econometric estimation methods of instrumental variables, generalised method of moments, and maximum likelihood. These core estimators are used extensively throughout the second half of the book and knowledge of them is a key asset in working through the later material. Also included in Part II are methods that deal with panel data and models with latent factors. A second course in financial econometrics might usefully begin with these five chapters, taking Part I as a given foundation.

Part III – Topics – introduces a number of special topics in financial economet-
rics, covering models of volatility, financial market microstructure, the econometrics of options, and methods relating to extreme values and copulas. One of the dominant features of financial time series is their volatility. Financial theory and empirical experience both demonstrate that there is often much less to explain in the levels of financial returns than there is to explain in their variation. Accordingly, three chapters of the book are devoted to modelling volatility. These chapters treat parametric univariate and multivariate models of volatility and introduce the more recent nonparametric modelling approach that is based on market realised volatility measured using high frequency data.

As in any project of this nature, sacrifices were made to keep the length of the book manageable. Some topics, for instance, are treated by example and illustration within a chapter rather than by devoting an entire chapter to their development. As a result the book is rich in real world examples drawn from financial markets for stocks, fixed income securities, exchange rates, derivatives, and real estate. As such, the coverage is intended to be extensive while not treating every topic in the same depth.

**Computation**

A fundamental principle guiding the inclusion of material in this book is whether the methods are available for easy implementation. In consequence, all results reported in the book may be reproduced using existing software packages like Stata and EViews. This choice is intended to enhance the usefulness of the material for beginning students. In some cases the programming languages in these packages need to be used to achieve full implementation of the illustrations. Of course, for those who actively choose to learn by programming themselves, the results are also reproducible in any of the common matrix programming languages. The numerical findings reported in the book are primarily rounded versions of the results generated using Stata.

The data files are all available for download from the book’s companion website (https://global.oup.com/academic/instructors/finects) in Stata format (.dta), EViews format (.wfl), comma delimited files (.csv), and as Excel spreadsheets (.xlsx).

**Acknowledgements**

Many colleagues, students, and research assistants have read and commented on parts of the book and in some cases even taught from early versions of the book. We are particularly grateful to Ahmad Bahir, Kit Baum, Ye Chen, Xiaohong Chen, Jieyang Chong, Adam Clements, Fulvio Corsi, Mark Doolan,
Renée Fry-McKibbin, Zhuo Huang, Marko Krause, Bei Luo, Cheng Liu, Andrew Patton, Shuping Shi, Daniel Smith, Chrismin Tang, Timo Teräsvirta, Stephen Thiele, Tomasz Woźniak and Lina Xu. A special thank you goes to Annastiina Silvennoinen and Glen Wade who were relentless in picking up typographical errors and factual inconsistencies, as well as suggesting alternative ways of presenting material. All remaining errors are the responsibility of the authors.

Stan Hurn, Vance L. Martin, Peter C. B. Phillips and Jun Yu
November 2018
## Contents

### I Fundamentals

1 Financial Asset Prices and Returns
   1.1 What is Financial Econometrics? ........................................ 3  
   1.2 Financial Assets ............................................................ 5  
   1.3 Equity Prices and Returns ............................................... 7  
   1.4 Stock Market Indices .................................................... 18  
   1.5 Bond Yields ............................................................... 23  
   1.6 Exercises ................................................................. 25  

2 Properties of Financial Data
   2.1 A First Look at the Data ............................................... 30  
   2.2 Summary Statistics ...................................................... 41  
   2.3 Percentiles and Value at Risk ......................................... 45  
   2.4 The Efficient Market Hypothesis .................................... 48  
   2.5 Exercises ................................................................ 51  

3 Linear Regression Models
   3.1 The Capital Asset Pricing Model ....................................... 58  
   3.2 A Multi-factor CAPM ...................................................... 63  
   3.3 Properties of Ordinary Least Squares ............................... 67  
   3.4 Diagnostics ................................................................. 72  
   3.5 Measuring Portfolio Performance ...................................... 80  
   3.6 Minimum Variance Portfolios ......................................... 82  
   3.7 Event Analysis ............................................................. 85
3.8 Exercises ................................................................. 87

4 Dynamic Modelling with Stationary Variables ........................................... 95
4.1 Stationarity .............................................................. 96
4.2 Univariate Time Series Models ........................................ 98
4.3 Autocorrelation and Partial Autocorrelations ...................... 103
4.4 Mean Aversion and Reversion in Returns .............................. 107
4.5 Vector Autoregressive Models ....................................... 109
4.6 Analysing VARs ....................................................... 115
4.7 Modelling Interest Rate Movements ................................ 119
4.8 Diebold-Yilmaz Spillover Index ..................................... 122
4.9 Exercises ................................................................. 124

5 Nonstationarity in Financial Time Series ............................................... 129
5.1 The Random Walk with Drift ........................................ 130
5.2 Characteristics of Financial Data .................................. 133
5.3 Dickey-Fuller Methods and Unit Root Testing ..................... 136
5.4 Beyond the Simple Unit Root Framework .......................... 143
5.5 Asset Price Bubbles ................................................... 148
5.6 Exercises ................................................................. 154

6 Cointegration ............................................................... 159
6.1 The Present Value Model and Cointegration ....................... 160
6.2 Vector Error Correction Models .................................... 165
6.3 Estimation ............................................................... 172
6.4 Cointegration Testing .................................................. 178
6.5 Parameter Testing ...................................................... 184
6.6 Cointegration and the Gordon Model ............................... 186
6.7 Cointegration and the Yield Curve .................................. 191
6.8 Exercises ................................................................. 196

7 Forecasting ................................................................. 203
7.1 Types of Forecasts ..................................................... 204
7.2 Forecasting Univariate Time Series Models ....................... 206
CONTENTS

7.3 Forecasting Multivariate Time Series Models .......................... 211
7.4 Combining Forecasts .................................................. 215
7.5 Forecast Evaluation Statistics ....................................... 218
7.6 Evaluating the Density of Forecast Errors ......................... 222
7.7 Regression Model Forecasts ......................................... 226
7.8 Predicting the Equity Premium ...................................... 229
7.9 Stochastic Simulation of Value at Risk ......................... 235
7.10 Exercises ............................................................... 238

II Methods ................................................................. 245

8 Instrumental Variables .................................................. 247
  8.1 Introduction ........................................................... 248
  8.2 Estimating the Risk-Return Tradeoff .............................. 249
  8.3 The General Instrumental Variables Estimator ............... 253
  8.4 Testing for Endogeneity ............................................ 257
  8.5 Weak Instruments .................................................. 259
  8.6 Consumption CAPM ............................................... 264
  8.7 Endogeneity and Corporate Finance ............................. 267
  8.8 Exercises ............................................................... 270

9 Generalised Method of Moments .................................... 275
  9.1 Single Parameter Models ........................................... 276
  9.2 Multiple Parameter Models ....................................... 279
  9.3 Over-Identified Models ............................................. 286
  9.4 Properties of the GMM Estimator ............................... 297
  9.5 Testing ................................................................. 304
  9.6 Consumption CAPM Revisited .................................. 307
  9.7 The CKLS Model of Interest Rates .............................. 311
  9.8 Exercises ............................................................... 314

10 Maximum Likelihood .................................................. 323
  10.1 Distributions in Finance ........................................... 324
10.2 Estimation by Maximum Likelihood .......................... 331
10.3 Applications ................................................. 334
10.4 Numerical Methods .......................................... 342
10.5 Properties .................................................. 344
10.6 Quasi Maximum Likelihood Estimation .................... 347
10.7 Testing ...................................................... 350
10.8 Exercises .................................................. 357

11 Panel Data Models ........................................... 365
  11.1 Types of Panel Data ...................................... 366
  11.2 Reasons for Using Panel Data ............................ 368
  11.3 Two Introductory Panel Models .......................... 372
  11.4 Fixed and Random Effects Panel Models ................. 377
  11.5 Dynamic Panel Models .................................. 386
  11.6 Nonstationary Panel Models ............................. 394
  11.7 Exercises .................................................. 400

12 Latent Factor Models ......................................... 409
  12.1 Motivation .................................................. 410
  12.2 Principal Components .................................... 412
  12.3 A Latent Multi-factor CAPM ............................. 420
  12.4 Dynamic Factor Models: the Kalman Filter ............. 423
  12.5 A Parametric Approach to Factors ....................... 431
  12.6 Stochastic Volatility ..................................... 435
  12.7 Exercises .................................................. 436

III Topics .................................................................. 445

13 Univariate GARCH Models .................................. 447
  13.1 Volatility Clustering ...................................... 448
  13.2 The GARCH Model ........................................ 451
  13.3 Asymmetric Volatility Effects ......................... 458
  13.4 Forecasting .................................................. 460
## CONTENTS

13.5 The Risk-Return Tradeoff ........................................ 465
13.6 Heatwaves and Meteor Showers ................................. 467
13.7 Exercises ..................................................... 470

### 14 Multivariate GARCH Models 479

14.1 Motivation .................................................... 480
14.2 Early Covariance Estimators ................................. 482
14.3 The BEKK Model ............................................. 483
14.4 The DCC Model .............................................. 490
14.5 Optimal Hedge Ratios ....................................... 496
14.6 Capital Ratios and Financial Crises .......................... 498
14.7 Exercises ..................................................... 501

### 15 Realised Variance and Covariance 509

15.1 High Frequency Data ........................................... 511
15.2 Realised Variance ............................................. 513
15.3 Integrated Variance ........................................... 516
15.4 Microstructure Noise ......................................... 519
15.5 Bipower Variation and Jumps ................................ 522
15.6 Forecasting .................................................... 525
15.7 The Realised GARCH Model ................................ 528
15.8 Realised Covariance .......................................... 531
15.9 Exercises ..................................................... 536

### 16 Microstructure Models 541

16.1 Characteristics of High Frequency Data ..................... 541
16.2 Limit Order Book .............................................. 542
16.3 Bid Ask Bounce ............................................... 544
16.4 Information Content of Trades ............................... 546
16.5 Modelling Price Movements in Trades ....................... 550
16.6 Modelling Durations ......................................... 556
16.7 Modelling Volatility in Transactions Time .................. 559
16.8 Exercises ..................................................... 562
CONTENTS

17 Options
17.1 Option Pricing Basics ................................................. 572
17.2 The Black-Scholes Option Price Model ........................... 575
17.3 A First Look at Options Data ...................................... 580
17.4 Estimating the Black-Scholes Model ......................... 581
17.5 Testing the Black-Scholes Model ................................ 587
17.6 Extensions to the Black-Scholes Model ....................... 590
17.7 Nonlinear Option Pricing ........................................ 593
17.8 Using Options to Estimate GARCH Models ............... 594
17.9 Exercises ......................................................... 597

18 Extreme Values and Copulas
18.1 Motivation .......................................................... 606
18.2 Evidence of Heavy Tails .......................................... 608
18.3 Extreme Value Theory ........................................... 611
18.4 Modelling Dependence using Copulas ..................... 617
18.5 Properties of Copulas ............................................. 620
18.6 Estimating Copula Models ..................................... 627
18.7 An MGARCH Model Using Copulas ....................... 630
18.8 Exercises ......................................................... 633

19 Conclusion .............................................................. 639

A Mathematical Preliminaries
A.1 Summation Notation ............................................... 641
A.2 Expectations Operator ............................................ 644
A.3 Differentiation .................................................... 645
A.4 Taylor Series Expansions ...................................... 647
A.5 Matrix Algebra ................................................... 652
A.6 Transposition of a Matrix ..................................... 656
A.7 Symmetric Matrix ................................................ 657

B Properties of Estimators
B.1 Finite Sample Properties ....................................... 667
List of Figures

1.1 Size of global derivatives market .................................................. 7
1.2 Stock market indices ................................................................. 19
1.3 Illustrating weighting schemes .................................................... 23
1.4 United States zero coupon yield curves ........................................ 24

2.1 Monthly United States equity price index from 1950 to 2016 ........ 31
2.2 Logarithm of monthly United States equity price index from 
   1950 to 1990 ........................................................................... 32
2.3 Monthly United States equity returns from 1950 to 1990 ......... 33
2.4 Monthly United States equity prices and dividends 1950 to 1990 34
2.5 Monthly United States dividends yield 1950 to 1990 ............. 35
2.6 Monthly United States zero coupon yields from 1946 to 1987 ... 36
2.7 United States zero coupon 9 month spread from 1950 to 1990 ... 37
2.8 Histogram of $/$ exchange rate returns ................................. 38
2.9 Histogram of durations between trades for AMR ................... 40
2.10 United States equity prices for the period 1950 to 2016 with 
     sample mean superimposed ....................................................... 42
2.11 Histogram of monthly United States equity returns 1950-2016 . 43
2.12 Daily 1% value at risk for Bank of America .......................... 47

3.1 Illustrating the ordinary least squares estimator of $\beta$ ............ 61
3.2 Fama-French and momentum factors ........................................ 66
3.3 Consistency of ordinary least squares ....................................... 69
3.4 Asymptotic normality of ordinary least squares ....................... 71
3.5 Illustrating autocorrelation in squared residuals ..................... 79
3.6 Microsoft and Walmart returns ........................................... 85
4.1 Equity price index 1871-2016 ............................................ 97
4.2 S&P 500 log returns 1871-2016 ............................................ 97
4.3 S&P 500 dividend returns 1871-2016 ................................. 107
4.4 VAR impulse responses for equity-dividend model .................. 117
4.5 Impulse responses for term structure model ......................... 122
5.1 Simulated random walk with drift ...................................... 132
5.2 Different filters applied to United States equity prices ............ 135
5.3 Simulated Dickey-Fuller distribution .................................... 139
5.4 Simulated distribution of Dickey-Fuller t-test ....................... 140
5.5 NASDAQ Index 1973 - 2009 ............................................. 149
5.6 Recursive estimation of ADF tests on the NASDAQ ............... 152
5.7 Rolling window estimation of ADF tests on the NASDAQ ........ 153
6.1 Logarithm of United States equity prices, dividends and earnings 161
6.2 Scatter plot of United States equity prices, dividends and earnings 163
6.3 Phase diagram to demonstrate equilibrium adjustment ............ 164
6.4 Residuals from cointegrating regression ............................... 180
6.5 3-D scatter plot of price, earnings and dividends ................... 187
6.6 Scatter plots of zero coupon yields ..................................... 192
6.7 Impulse responses for term structure VECM ......................... 196
7.1 Ex post forecasts of United States equity returns ................... 220
7.2 Probability integral transform .......................................... 224
7.3 Illustrating the probability integral transform ....................... 225
7.4 Illustrating the probability integral transform ....................... 227
7.5 Equity premium, dividend yield and dividend price ratio ........ 230
7.6 Recursive coefficients from predictive regressions ................ 232
7.7 Evaluating predictive regressions of the equity premium .......... 233
7.8 Stochastic simulation of equity prices .................................. 237
7.9 Simulating value at risk ................................................. 237
LIST OF FIGURES

8.1 Risk-return tradeoff ................................................. 251
8.2 Weighted instrument for extended CAPM model ............... 255
8.3 Illustrating the effect of weak instruments .................... 261

9.1 Moment condition for present value model .................... 277
9.2 Durations data ...................................................... 279
9.3 Moment condition for the durations model .................... 280
9.4 Excess returns to Exxon and S&P 500 ......................... 282
9.5 Moment conditions for the CAPM ............................ 283
9.6 Gradient of over-identified duration model ................ 295
9.7 Consistency of GMM estimator ............................... 300
9.8 Asymptotic normality of GMM estimator .................. 302
9.9 Monthly United States zero coupon bond yields 1946 to 1991 . 312

10.1 Distributions of financial returns ............................. 326
10.2 Distributions of financial returns ............................. 326
10.3 Log-normal distribution for equity prices ................... 328
10.4 Exponential model ............................................... 330
10.5 The log likelihood function of exponential model ........ 335
10.6 Transitional density of Eurodollar interest rates .......... 341

11.1 Random versus fixed effects ..................................... 384
11.2 Panel scatter plots of the present value relationship ...... 395

12.1 Yields on United States Treasury bonds of various maturities . 411
12.2 Estimated factors for yields on United States Treasuries ..... 419
12.3 Comparison of actual and latent market factor ............. 423
12.4 Latent factor for the term structure .......................... 431
12.5 Diebold and Li (2006) factor loadings ....................... 432
12.6 Diebold and Li (2006) factors ................................ 433
12.7 Forecast yield curves ............................................. 434
12.8 £/$ exchange rate ................................................... 436
12.9 Stochastic volatility of £/$ exchange rate .................. 437
LIST OF FIGURES

13.1 Daily returns to stock market indices .................................. 449
13.2 Distributions of returns to the DAX ................................. 450
13.3 Volatility in turbulent and tranquil regimes ......................... 451
13.4 News Impact Curves ...................................................... 461
13.5 GARCH(1,1) forecasts .................................................. 463
13.6 Illustrating forecast loss functions .................................. 465

14.1 Conditional variances and covariance of Microsoft and the S&P
    500 index .................................................................... 484
14.2 Time-varying \( \beta \)-risk for Microsoft ............................... 485
14.3 Time-varying beta from BEKK model ............................... 489
14.4 Returns to 4 industry portfolios ..................................... 494
14.5 Dynamic hedge ratios .................................................... 497

15.1 IBM tick data .................................................................. 512
15.2 Signature plot for high frequency IBM data ....................... 515
15.3 Stock price dynamics ...................................................... 517
15.4 Stock price dynamics with jumps .................................... 523
15.5 Returns and realised variance for IBM ............................ 527
15.6 Realised variance and covariance signature plots ............... 535

16.1 Autocorrelations for 5 minute BHP returns ....................... 545
16.2 Quote revision process for BHP ....................................... 548
16.3 Distribution of changes in bid and ask prices .................... 550
16.4 Durations between trades for BHP ................................. 557
16.5 Diurnal pattern of durations for BHP ............................. 560

17.1 Option pricing simulation ................................................. 573
17.2 Plots of the forecast distribution for asset price ............... 576
17.3 S&P 500 index and returns ............................................ 582
17.4 Volatility smile .............................................................. 590
17.5 Conditional volatility of log returns on the S&P 500 index .... 592

18.1 Daily S&P 500 percentage returns ................................. 607
18.2 Daily S&P 500 percentage returns ................................. 610
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>18.3</td>
<td>Extreme value distributions</td>
<td>613</td>
</tr>
<tr>
<td>18.4</td>
<td>Tail index plot</td>
<td>615</td>
</tr>
<tr>
<td>18.5</td>
<td>Scatter plot of S&amp;P indices</td>
<td>619</td>
</tr>
<tr>
<td>18.6</td>
<td>Simulating data from copulas</td>
<td>621</td>
</tr>
<tr>
<td>B.1</td>
<td>Consistency property of the sample mean</td>
<td>671</td>
</tr>
<tr>
<td>B.2</td>
<td>Asymptotic normality property of the sample mean</td>
<td>672</td>
</tr>
</tbody>
</table>
Part I

Fundamentals
Chapter 1

Financial Asset Prices and Returns

1.1 What is Financial Econometrics?

No simple definition is sufficient to answer this question, as many past writers have pointed out. The subject is best described through its many different activities. These include the formulation of financial models intended for empirical implementation, methods of estimation and inference with these models, and their use for forecasting, for policy analysis, and for understanding financial phenomena.

As a subject, financial econometrics is interdisciplinary. It draws on ideas and methods from finance, economics, probability, statistics and applied mathematics, while at the same time providing a rich source bed of new ideas for modelling, estimation, and inference. Partly because of this diversity, financial econometrics is a vast and growing discipline with applications that stretch across the social and business sciences. Its primary tasks stem from the particular nature of financial data and the body of financial theory that has been developed to explain the complex world of finance and financial instruments that surrounds us.

While the origins of financial econometrics may be traced back to early empirical studies of stock prices, bond yields and interest rates, the subject began to take aspects of its modern form during the 1980s. At this time, the methods of time series econometrics evolved in ways that were especially beneficial to studying financial data, taking account of features such as the wandering nature of financial asset prices, the volatility of financial returns, and the availability of ultra-high frequency observations. These new modelling and inferential tools drawn from time series econometrics and other constituent disci-
plines joined with a growing specialisation amongst econometricians working with financial data to promote the development of a new discipline with the common goal of searching for a deeper understanding of the way in which financial markets work and financial asset prices are determined. Out of this understanding and sustained research, it is to be hoped, regulators and policy makers will be better equipped to assist in monitoring markets towards the lofty goal of financial stability and to guide the smooth functioning of financial markets in the face of crisis.

Central to the success of this scientific process is the initial step of establishing a reliable data set that is well-suited to the various tasks of econometric investigation. The financial data of primary interest in applications are the prices of financial assets and the yields or returns to investments from those assets. The first logical step in the study of financial econometrics, therefore, is to become familiar with the many different types of financial assets, how prices for these assets are quoted and reported, and how yields or returns to investment in such assets are constructed.

A distinguishing feature of financial econometrics that sets the subject apart from many other applications of econometrics, particularly macroeconomics, is the abundance of financial data. Terabytes of data covering a vast array of financial assets are created daily, producing high dimensional data sets that carry fine-grain transaction level details recording the continuous pulse of financial markets. These data are used in the industry by financial firms searching for investment opportunities and they provide an extraordinary digitised resource for financial econometricians. Unlike macroeconomics, in finance there is no paucity of data for testing hypotheses of interest. But superabundance of observations is no testament to quality or the absence of measurement error, missing observations, data revisions, or subtleties associated with transaction-level data.

Much work is frequently needed to get financial data into the clean form that is necessary for meaningful empirical analysis. These problems go well beyond the superficial and they can involve deep questions associated with the very structure of trading in financial markets. Addressing the plethora of financial data, the risks associated with data-mining, the subtleties involved in transaction level observations, and the probabilistic foundations of modelling and inference with such data have enfranchised the energies of the large and growing community of scholars in financial econometrics.

This chapter does not attempt to cover all the interesting twists and turns of data-creation in the financial world that applied researchers in financial econometrics have to face in their empirical work. But it will highlight some of these issues and stimulate a renewed awareness of the famous adage that empirical results are only as good as the data on which they are based.
1.2 Financial Assets

Although they may have no intrinsic physical worth, financial assets derive value from the contractual claims they place on a stream of services or cash flows. The major categories of financial assets that will be used in this book are cash, fixed-income securities, equity securities and derivative securities.

Cash
Cash represents a claim on the stream of services that it can secure by virtue of its role as a medium of exchange. One particularly important financial transaction that may be regarded as a cash investment is dealing in foreign exchange. The exchange rate is simply the price of one currency in terms of another. So trading in currencies may be regarded as investments in cash.

Fixed-Income Securities
Fixed-income securities provide two sources of return. The first corresponds to a stream of interest payments (or coupons) that are made at fixed, regular intervals, and the second is to the eventual return of principal at maturity. Although the original distinguishing feature of this class of financial asset was that the periodic payment was known in advance, recent developments in financial markets link many of these payments to a particular short-term interest rate and some are even linked to the prevailing inflation rate.

Money market fixed-income securities are short-term assets whose markets are particularly active (or liquid). There is now a bewildering array of money market instruments available to study. But only two will feature in this book.

- Treasury Bills are the simplest form of government debt. The government sells Treasury Bills in the money market and redeems them at the maturity of the bill. No interest is payable during the life of the bill and so they trade at discount to the face value of the bill that will be paid at maturity. The most common maturities are 3, 6 and 9 months.

- Eurodollar Deposits are the deposits of United States banks which are denominated in US$ but held with banks outside the United States. Most of these deposits have a relatively short maturity (less than 6 months) and the Eurodollar deposit rate is commonly used as a representative short-term interest rate.

The bond market is the place where longer term borrowing of governments or corporations is conducted. A bond is a security which promises to pay the owner of the bond its face value at the time of the maturity of the bond and usually an ongoing coupon payment prior to maturity. There are also zero-coupon bonds that pay no regular interest and are therefore traded at prices that are below their face value. In recent times this distinction has become less important because zero-coupon bonds may be created from coupon paying
bonds by separating the coupons from principal and trading each of these components independently. This process is known as stripping.

Another common way in which the fixed-income securities market is classified is by the issuer of the securities. For instance, a distinction is sometimes made between bonds issued by financial intermediaries (FI bonds) and non-financial intermediaries (NFI bonds). Financial intermediaries are entities that facilitate financial transactions between two or more parties and include commercial banks, investment banks and insurance companies.

Equity Securities

Equities or common stocks give the owner an equity stake in a company and a corresponding claim on company assets and earnings. Equities can be bought and sold on stock markets. Stocks give the owner the right to a payment which represents the distribution of some of the company’s earnings, which is known as a dividend. The dividend is usually expressed as the amount each share receives or as a percentage of the current market price, which is referred to as the dividend yield.

Derivative Securities

Derivative securities provide a payoff based on the value(s) of other assets such as commodities, bonds, or stocks. Such securities therefore derive their own value from the market performance of the other underlying assets to which they are attached. Derivatives started out as over-the-counter (OTC) trades where interested parties made mutually beneficial trades. In recent years as more standardised contracts have emerged, derivatives have been very actively traded on exchanges such as the Chicago Board of Exchange.

Two classes of derivative securities are emphasised in this book.

- Options contracts offer the buyer the right, but not the obligation, to buy (call option) or sell (put option) some designated financial asset (the underlying asset from which the option derives its value) at a particular price during a certain period of time or on a specific date.

- Futures contracts specify the delivery of either an asset or a cash value at a time known as the maturity for an agreed price which is payable at maturity. The entity who commits to purchase the asset on delivery takes a long position. The entity who commits to delivering the asset takes a short position.

One of the most significant developments in financial markets in recent years has been the growth of derivatives markets as illustrated in Figure 1.1. The problem with measuring the size of the derivatives market stems from the fact that there is a large volume of OTC trades which make it difficult to quantify the exact volume of derivative trading. What is clear from Figure 1.1, however, is that the combined outstanding value of derivatives is several or-
1.3. EQUITY PRICES AND RETURNS

In this section, the prices of financial assets and the returns to holding these assets will be couched in terms of common stocks. Stocks represent an equity claim on the company and typically, although not always, receive a regular stream of dividend payments. Prices and returns associated with other financial assets are determined in a similar way.

1.3.1 Prices

The most basic data in financial econometrics are the prices of financial assets. The price of an equity security is defined in terms of the dollar (or other currency denomination) amount at which a transaction can occur (a quoted price) or has occurred (an historical transaction price). When dealing with high-frequency data the appropriate prices are usually quoted prices. An illustration is provided in Table 1.1 which gives quoted prices obtained from Yahoo Finance for common stock in the United States company Boeing on 12 September 2014.
CHAPTER 1. FINANCIAL ASSET PRICES AND RETURNS

Table 1.1

Quoted prices on Yahoo Finance for The Boeing Company (BA):
12 September 2014.

<table>
<thead>
<tr>
<th>The Boeing Company (BA) - NYSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>127.64 ↓ 0.58(0.45%)</td>
</tr>
<tr>
<td>Prev Close: 128.22</td>
</tr>
<tr>
<td>Open: 127.82</td>
</tr>
<tr>
<td>Bid: 127.50</td>
</tr>
<tr>
<td>Ask: 127.84</td>
</tr>
<tr>
<td>Day's Range: 127.20 - 127.99</td>
</tr>
<tr>
<td>52wk Range: 109.14 - 144.57</td>
</tr>
<tr>
<td>Volume: 1,988,616</td>
</tr>
<tr>
<td>Market Cap: 91,98 Bil.</td>
</tr>
</tbody>
</table>

Source: https://au.finance.yahoo.com

Recording a price for the purpose of econometric analysis is not as straightforward as it might seem. A number of alternatives are available. In addition to the previous day closing price and the current day opening price there are also prevailing bid and ask prices. The bid price is the maximum price that buyers are willing to pay for the stock and the ask price is the minimum price that sellers are willing to accept for the stock. The differential between the bid and ask price is called the spread. Many studies that use intra-day data (popularly known as high-frequency data) compromise this complexity by using the midpoint of the bid and ask prices as the best summary estimate of the prevailing current price. This convention simplifies data analysis but circumvents important details of the transaction price determination process. In doing so it produces intriguing econometric problems arising from the impact of the neglected market microstructure in modelling prices, as we now explain.

The practice of using summary estimates of prices, like the midpoint of the bid-ask spread, points to an unusual feature of high frequency data that affects econometric work: increasing the number of observations by using intra-day data need not always increase efficiency or improve understanding. More data often means that there is more to explain. Indeed, in the present case, adding more data by using more frequent observations changes the focus of attention towards a microscopic focus on the transaction process itself - what is called market microstructure. This new focus in turn raises the dimensionality of the econometric problem by virtue of the complexity of the transaction process which brings buyers and sellers together through a market determination process that involves multiple bid-ask order layers in which random elements may enter the price determination process in the microcosm of market forces.

When dealing with historical prices recorded at lower frequencies the situation is less complex. Table 1.2 reports the historical daily prices for the United States stock Microsoft for the month of August 2014. The choice for the researcher is now simpler. We have the recorded opening price, closing price, daily highs and lows, and the adjusted closing price. In most cases it is con-
1.3. EQUITY PRICES AND RETURNS

Convenient to choose the adjusted closing price (denoted by \( \text{Close}^* \)), which is adjusted for stock splits and dividends.

Table 1.2

Daily prices for the United States stock Microsoft (MSFT) for the month of August 2014. All prices are quoted in US$. The column, \( \text{Close}^* \), gives the closing price adjusted for dividends and stock splits. A dividend of US$ 0.28 per share was paid on 19 August 2014.

<table>
<thead>
<tr>
<th>Date</th>
<th>Open</th>
<th>High</th>
<th>Low</th>
<th>Close</th>
<th>Volume</th>
<th>( \text{Close}^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 Aug 14</td>
<td>45.09</td>
<td>45.44</td>
<td>44.86</td>
<td>45.43</td>
<td>21607600</td>
<td>45.43</td>
</tr>
<tr>
<td>28 Aug 14</td>
<td>44.75</td>
<td>44.98</td>
<td>44.61</td>
<td>44.88</td>
<td>17657600</td>
<td>44.88</td>
</tr>
<tr>
<td>27 Aug 14</td>
<td>44.90</td>
<td>45.00</td>
<td>44.76</td>
<td>44.87</td>
<td>20823000</td>
<td>44.87</td>
</tr>
<tr>
<td>26 Aug 14</td>
<td>45.31</td>
<td>45.40</td>
<td>44.94</td>
<td>45.01</td>
<td>14873100</td>
<td>45.01</td>
</tr>
<tr>
<td>25 Aug 14</td>
<td>45.40</td>
<td>45.44</td>
<td>45.04</td>
<td>45.17</td>
<td>16898100</td>
<td>45.17</td>
</tr>
<tr>
<td>22 Aug 14</td>
<td>45.35</td>
<td>45.47</td>
<td>45.07</td>
<td>45.15</td>
<td>18294500</td>
<td>45.15</td>
</tr>
<tr>
<td>21 Aug 14</td>
<td>44.84</td>
<td>45.25</td>
<td>44.83</td>
<td>45.22</td>
<td>22272000</td>
<td>45.22</td>
</tr>
<tr>
<td>20 Aug 14</td>
<td>45.34</td>
<td>45.40</td>
<td>44.90</td>
<td>44.95</td>
<td>24750700</td>
<td>44.95</td>
</tr>
<tr>
<td>19 Aug 14</td>
<td>44.97</td>
<td>45.34</td>
<td>44.83</td>
<td>45.33</td>
<td>28115600</td>
<td>45.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dividend US$ 0.28</th>
</tr>
</thead>
<tbody>
<tr>
<td>18 Aug 14</td>
</tr>
<tr>
<td>15 Aug 14</td>
</tr>
<tr>
<td>14 Aug 14</td>
</tr>
<tr>
<td>13 Aug 14</td>
</tr>
<tr>
<td>12 Aug 14</td>
</tr>
<tr>
<td>11 Aug 14</td>
</tr>
<tr>
<td>8 Aug 14</td>
</tr>
<tr>
<td>7 Aug 14</td>
</tr>
<tr>
<td>6 Aug 14</td>
</tr>
<tr>
<td>5 Aug 14</td>
</tr>
<tr>
<td>4 Aug 14</td>
</tr>
<tr>
<td>1 Aug 14</td>
</tr>
</tbody>
</table>

Source: https://au.finance.yahoo.com

The effect of a dividend payment is to lower the price by the amount of the dividend so that the closing price on 18 August is greater than the opening price on 19 August. In order to ensure that the effect of the dividend is smoothed out in historical prices, the correction is to subtract the dividend from the closing price on the previous day, compute the quotient \( \frac{P_{t-1} - D_t}{P_{t-1}} \), and then multiply all previous prices by this factor. On 18 August the closing price and the adjustment factor are given by

\[
$44.83 = 45.11 - 0.28 \quad \text{and} \quad \frac{45.11 - 0.28}{45.11} = 0.9938,
\]

respectively. In consequence, the adjusted closing price on 15 August is

\[
$44.51 = 44.79 \times 0.9938.
\]
Note that this process of adjustment means that the historical prices do not necessarily reflect the actual prices at which trades took place.

The adjustment process for a stock split is similar. Suppose, for instance, that a 2-for-1 stock split occurs in which a company replaces each existing share by two shares (or some other multiple). Then the price of an individual share is immediately halved (or scaled by the otherwise appropriate fraction). Such splits make shares appear more affordable even though the underlying market capitalization of the company has not changed. To avoid the artificially induced discontinuity in the share price at the time of the split, all historical prices need to be divided by 2 and the historical volume series correspondingly multiplied by 2 so that the price after the split and the price before the split are comparable.

A further issue of data comparability is the presence of non-trading days in the raw data. For instance, a close look at the calendar days in the first column of Table 1.2 reveals a number of missing days, each of which corresponds (in this instance) to weekends and public holidays. But there may be days other than public holidays and weekends when a stock does not trade. In addition, when comparing time series of stock prices from different countries, public holidays do not always fall on the same days. In preparing data for empirical work all these details need attention.

1.3.2 Returns

The return to a financial asset probably receives more attention in financial econometrics than does the price of an asset, although the movement of stock prices over long historical periods is also of substantial interest to investors and is relevant in practical econometric work dealing with long run trends.

Broadly speaking a financial return provides a measure of outcome of the decision to invest in a financial asset. This measure accounts not only for the capital gain or loss due to the price change over the holding period of the asset but also for the cumulative impact of the contractual stream of cash flows that take place over the course of the holding period.

In principle, a financial asset might be held for an indeterminate period. Historically, stock prices were usually measured at daily, weekly, and monthly frequencies. In that case, the holding period of the investment is limited to a multiple of this frequency. But with the advent of readily available high-frequency data, returns can be computed for most holding periods, even extremely short ones. The latter have become much more important with computerized trading practice.
1.3. **EQUITY PRICES AND RETURNS**

**Dollar Returns**

The simplest measure of return on holding an asset for \( k \) periods between time \( t \) and \( t - k \) is the dollar return, denoted \( \$R_{kt} \), given by the price differential over this period

\[
\$R_{kt} = P_t - P_{t-k}.
\]

Although this measure is a simple intuitive response to the problem of computing the return to an investment its major drawback is that it is not scale-free and does not measure the return relative to the initial investment. Moreover, this measure depends on the unit in which prices (and dividends) are quoted. To make returns comparable across assets and across international financial markets scale-free measures of returns are required.

**Simple Returns**

The simple return on an asset between time \( t - 1 \) and \( t \) is given by

\[
R_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1. \tag{1.1}
\]

The relative price ratio \( P_t / P_{t-1} \), also known as the price relative (or prl for short), is a useful quantity to compute. If the ratio is greater than 1 then returns are positive and if it is less than 1 returns are negative. Equation 1.1 may be rearranged as

\[
1 + R_t = \frac{P_t}{P_{t-1}},
\]

in which \( 1 + R_t \) is known as the simple gross return. The usefulness of the simple gross return is that it represents the value at time \( t \) of investing $1 at time \( t - 1 \).

The return to holding an asset for \( k \) periods, \( R_t(k) \), is given by

\[
R_t(k) = \frac{P_t}{P_{t-k}} - 1
\]

\[
= \frac{P_t}{P_{t-1}} \times \frac{P_{t-1}}{P_{t-2}} \times \cdots \times \frac{P_{t-k+2}}{P_{t-k+1}} \times \frac{P_{t-k+1}}{P_{t-k}} - 1
\]

\[
= (1 + R_t) \times (1 + R_{t-1}) \times \cdots \times (1 + R_{t-k+2}) \times (1 + R_{t-k+1}) - 1
\]

\[
= \prod_{j=0}^{k-1} (1 + R_{t-j}) - 1. \tag{1.2}
\]

The point to be emphasised in this calculation is that simple returns are not additive when computing multi-period returns because of the multiplicative effect of period-by-period returns.
CHAPTER 1. FINANCIAL ASSET PRICES AND RETURNS

If the data frequency is monthly, then the simple return for a holding period of one year is given by

\[ R_t(12) = \left[ \prod_{j=0}^{11} (1 + R_{t-j}) \right] - 1. \tag{1.3} \]

The most common period over which a return is quoted is one year and returns data are commonly presented in per annum terms. This means that the current monthly return needs to be appropriately scaled so that it is interpretable as an annual return and expressed on a per annum basis. In the case of monthly returns, the associated annualised simple return is computed as

\[ \text{Annualised } R_t(12) = (1 + R_t)^{12} - 1. \tag{1.4} \]

Expression (1.4) is obtained from (1.3) by making the assumption that the best guess of the per annum return is that the current monthly return will persist for the next 12 months. In this case, all the terms in the product expansion (in square brackets) of equation (1.3) will be identical.

Log Returns

The log return of an asset is defined as

\[ r_t = \log(1 + R_t) = \log P_t - \log P_{t-1}. \tag{1.5} \]

Log returns are also referred to as continuously compounded returns. To understand why this is so it is convenient to use the exponential constant \( e \). The Swiss mathematician Leonhard Euler (1707-1783) named this constant, introduced the letter \( e \) to represent it, showed its now well-known exponential series representation, and proved its form in terms of the limiting operation

\[ e \equiv \lim_{s \to \infty} \left( 1 + \frac{1}{s} \right)^s \approx 2.71828. \]

Somewhat earlier in 1683, another Swiss mathematician Jacob Bernoulli (1655-1705) attempted to find this limit in studying the effect of continuously compounded interest. Its discovery is often attributed to him and links the mathematics of compound interest with the subjects of accounting, finance, and economics. The limit formula above represents the value of an account at the end of the year which started with $1.00 and paid 100% interest per year but with the interest compounded continuously over time rather than at discrete intervals during the year.

If \( m \) is the compounding period and \( r_t \) the return, then it follows from above that

\[ P_t = P_{t-1} \left( 1 + \frac{r_t}{m} \right)^m. \]
Continuous compounding is produced when \( m \to \infty \) leading to

\[
P_t = P_{t-1} \lim_{m \to \infty} \left( 1 + \frac{r_t}{m} \right)^m. \tag{1.6}
\]

Let \( s = m/r_t \) in this formula. Then the expression in (1.6) may be rewritten as

\[
P_t = P_{t-1} \lim_{s \to \infty} \left[ \left( 1 + \frac{1}{s} \right)^{s r_t} \right]
= P_{t-1} \left[ \lim_{s \to \infty} \left( 1 + \frac{1}{s} \right)^{s r_t} \right]
= P_{t-1} e^{r_t}. \tag{1.7}
\]

Taking logarithms of expression (1.7) yields the definition of the log returns given in equation (1.5).

Log returns are particularly useful because of the simplification they allow in dealing with multi-period returns. For example, the 2-period return is given by

\[
r_t(2) = \log P_t - \log P_{t-2}
= (\log P_t - \log P_{t-1}) + (\log P_{t-1} - \log P_{t-2})
= r_t + r_{t-1}. \tag{1.8}
\]

By extension the \( k \)-period return is

\[
r_t(k) = \log P_t - \log P_{t-k}
= (\log P_t - \log P_{t-1}) + (\log P_{t-1} - \log P_{t-2}) + \cdots + (\log P_{t-k+1} - \log P_{t-k})
= r_t + r_{t-1} + \cdots + r_{t-(k-1)}
= \sum_{j=0}^{k-1} r_{t-j}. \tag{1.9}
\]

In other words, the \( k \)-period log return is simply the sum of the single period log returns over the pertinent period.

For the case of data observed monthly, the annual log return is

\[
r_t(12) = \log P_t - \log P_{t-12} = \sum_{j=0}^{11} r_{t-j}. \tag{1.10}
\]

Once again, expression (1.9) may be used to obtain the returns expressed on a per annum basis by simply multiplying all monthly returns by 12, making the implicit assumption that the best guess of the per annum return is that the current monthly return will persist for the next 12 months.

By analogy, if prices are observed quarterly, then the individual quarterly returns can be annualised by multiplying the quarterly returns by 4. Similarly,
if prices are observed daily, then the daily returns are annualised by multiplying the daily returns by the number of trading days 252. The choice of 252 for the number of trading days in a calendar year is an approximation because of the effect of public holidays, leap years, and additional days of trading interruption. Other choices are 250 and, very rarely, the number of calendar days, 365, is used.

Table 1.3 provides calculations based on historical monthly prices for the United States stock Microsoft showing the mechanics of return computations from the price of a stock. Note that no return figures are reported for January 2012. Their absence emphasises that an observation is lost at the beginning of the sample when computing returns because the price of the stock before the start of the sample period is not available. The monthly dollar, simple and log returns to Microsoft for February 2012 are respectively

\[
R_t = \frac{31.74 - 29.53}{29.53} = 0.075 = 7.5%,
\]

\[
r_t = \log(1 + 0.075) = 0.072 = 7.2%.
\]

These calculations demonstrate that continuously compounded returns are very similar to simple returns as long as the return is relatively small, which it generally will be for monthly or daily returns. Indeed, it is only really at the third decimal place that the differences between the two definitions of returns become readily apparent.
1.3. 

**EQUITY PRICES AND RETURNS**

Table 1.3

Monthly prices for the United States stock Microsoft for the years 2012 and 2013. Also shown are alternative measures of the one-month return to holding Microsoft. Prices are month-end closing prices adjusted for splits and dividends quoted in US$.

<table>
<thead>
<tr>
<th>Date</th>
<th>Price</th>
<th>Prel</th>
<th>Monthly Dollar Return</th>
<th>Monthly Simple Return</th>
<th>Monthly Log Return</th>
<th>Annual Simple Return</th>
<th>Annual Log Return</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan 2012</td>
<td>29.530</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Feb 2012</td>
<td>31.740</td>
<td>1.075</td>
<td>2.210</td>
<td>0.075</td>
<td>0.072</td>
<td>1.378</td>
<td>0.866</td>
</tr>
<tr>
<td>Mar 2012</td>
<td>32.250</td>
<td>0.510</td>
<td>0.016</td>
<td>0.016</td>
<td>0.016</td>
<td>0.211</td>
<td>0.191</td>
</tr>
<tr>
<td>Apr 2012</td>
<td>32.020</td>
<td>0.993</td>
<td>-0.230</td>
<td>-0.007</td>
<td>-0.007</td>
<td>-0.082</td>
<td>-0.086</td>
</tr>
<tr>
<td>May 2012</td>
<td>29.190</td>
<td>0.912</td>
<td>-2.830</td>
<td>-0.088</td>
<td>-0.093</td>
<td>-0.671</td>
<td>-1.110</td>
</tr>
<tr>
<td>Jun 2012</td>
<td>29.470</td>
<td>1.048</td>
<td>1.400</td>
<td>0.048</td>
<td>0.047</td>
<td>0.754</td>
<td>0.562</td>
</tr>
<tr>
<td>Jul 2012</td>
<td>30.590</td>
<td>1.048</td>
<td>1.400</td>
<td>0.048</td>
<td>0.047</td>
<td>0.754</td>
<td>0.562</td>
</tr>
<tr>
<td>Aug 2012</td>
<td>30.820</td>
<td>1.048</td>
<td>1.400</td>
<td>0.048</td>
<td>0.047</td>
<td>0.754</td>
<td>0.562</td>
</tr>
<tr>
<td>Sep 2012</td>
<td>29.780</td>
<td>0.966</td>
<td>-1.040</td>
<td>-0.034</td>
<td>-0.034</td>
<td>-0.338</td>
<td>-0.412</td>
</tr>
<tr>
<td>Oct 2012</td>
<td>28.530</td>
<td>0.958</td>
<td>-1.245</td>
<td>-0.042</td>
<td>-0.043</td>
<td>-0.401</td>
<td>-0.512</td>
</tr>
<tr>
<td>Nov 2012</td>
<td>26.620</td>
<td>0.933</td>
<td>-1.915</td>
<td>-0.067</td>
<td>-0.069</td>
<td>-0.566</td>
<td>-0.834</td>
</tr>
<tr>
<td>Dec 2012</td>
<td>26.730</td>
<td>1.004</td>
<td>0.110</td>
<td>0.004</td>
<td>0.004</td>
<td>0.051</td>
<td>0.049</td>
</tr>
<tr>
<td>Jan 2013</td>
<td>27.470</td>
<td>1.028</td>
<td>0.740</td>
<td>0.028</td>
<td>0.027</td>
<td>0.388</td>
<td>0.328</td>
</tr>
<tr>
<td>Feb 2013</td>
<td>27.800</td>
<td>1.012</td>
<td>0.330</td>
<td>0.012</td>
<td>0.012</td>
<td>0.154</td>
<td>0.143</td>
</tr>
<tr>
<td>Mar 2013</td>
<td>28.610</td>
<td>1.029</td>
<td>0.810</td>
<td>0.029</td>
<td>0.029</td>
<td>0.411</td>
<td>0.345</td>
</tr>
<tr>
<td>Apr 2013</td>
<td>33.100</td>
<td>1.157</td>
<td>4.490</td>
<td>0.157</td>
<td>0.146</td>
<td>4.751</td>
<td>1.749</td>
</tr>
<tr>
<td>May 2013</td>
<td>34.880</td>
<td>1.054</td>
<td>1.780</td>
<td>0.054</td>
<td>0.052</td>
<td>0.875</td>
<td>0.629</td>
</tr>
<tr>
<td>Jun 2013</td>
<td>34.530</td>
<td>0.990</td>
<td>-0.350</td>
<td>-0.010</td>
<td>-0.010</td>
<td>-0.114</td>
<td>-0.121</td>
</tr>
<tr>
<td>Jul 2013</td>
<td>31.830</td>
<td>0.922</td>
<td>-2.700</td>
<td>-0.078</td>
<td>-0.081</td>
<td>-0.624</td>
<td>-0.977</td>
</tr>
<tr>
<td>Aug 2013</td>
<td>33.400</td>
<td>1.049</td>
<td>1.570</td>
<td>0.049</td>
<td>0.048</td>
<td>0.782</td>
<td>0.578</td>
</tr>
<tr>
<td>Sep 2013</td>
<td>33.310</td>
<td>1.097</td>
<td>-0.090</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.032</td>
<td>-0.032</td>
</tr>
<tr>
<td>Oct 2013</td>
<td>35.350</td>
<td>1.061</td>
<td>2.040</td>
<td>0.061</td>
<td>0.059</td>
<td>1.041</td>
<td>0.713</td>
</tr>
<tr>
<td>Nov 2013</td>
<td>38.130</td>
<td>1.079</td>
<td>2.780</td>
<td>0.079</td>
<td>0.076</td>
<td>1.480</td>
<td>0.908</td>
</tr>
<tr>
<td>Dec 2013</td>
<td>37.430</td>
<td>0.982</td>
<td>-0.700</td>
<td>-0.018</td>
<td>-0.019</td>
<td>-0.199</td>
<td>-0.222</td>
</tr>
</tbody>
</table>

Source: Bloomberg.

Despite the similarities in the two measures of returns, appreciable differences emerge when the returns are annualised. For the simple return in February 2012 the calculation is

\[ R_t(12) = (1 + 0.075)^{12} - 1 = 1.378 \times 100\% = 137.8\% . \]

By contrast the annualised log return is

\[ r_t(12) = 12 \times 0.072 = 0.866 = 86.6\% . \]

Note that the practice of quoting figures as annual rates is usually related to scaling the data. Returns, when computed over the time interval of a day or even shorter intervals, can be relatively small in value and this may lead to arithmetic errors when doing complex computations involving the returns. Annualising the return scales can help to alleviate this problem.
Dealing with Dividends

Adjusting the computation of returns for the payment of a dividend, $D_t$, between time $t - 1$ and $t$, is relatively straightforward. The dollar return becomes

$$R_t = P_t + D_t - P_{t-1},$$

in which $P_t$ and $P_{t-1}$ are the unadjusted prices. The simple and gross returns are then given by

$$R_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} + \frac{D_t}{P_{t-1}} - 1,$$

and

$$(1 + R_t) = \frac{P_t + D_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} + \frac{D_t}{P_{t-1}},$$

respectively. It is apparent from (1.11) and (1.12) that the simple and gross returns to a stock in the presence of a dividend payment are easily computed in terms of the price relative and the dividend yield.

Adjusting log returns for a dividend payment simply requires using the correct definition of gross simple returns when taking logarithms

$$r_t = \log(1 + R_t) = \log \left(1 + \frac{P_t + D_t - P_{t-1}}{P_{t-1}}\right) = \log \left(\frac{P_t}{P_{t-1}} + \frac{D_t}{P_{t-1}}\right).$$

Much of the earlier discussion concerning the computation of returns has reflected common practice and ignored the issue of dividends. This practice stems from the reality that dividends are paid relatively infrequently and constitute a minor proportion of the overall return compared with price movements.

Excess Returns

The difference between the return on a risky financial asset and a risk-free interest rate, denoted $r_{ft}$, is known as the excess return. The risk-free rate is often taken to be the interest rate on a government bond. The simple and log excess returns on an asset are therefore defined as $R_t - r_{ft}$ and $r_t - r_{ft}$, respectively. In computing the excess returns it is important to ensure that the risk-free interest rate is expressed in the same unit of time as the return on the risky financial asset. For example, interest rates are normally quoted as annual rates so in the case of monthly log returns the quoted annual risk-free interest rate would need to be divided by 12.

1.3.3 Portfolio Returns

Financial econometric work is often concerned not with the return to a single asset as the prime object of the investigation but rather the return to a portfolio of financial assets. Attention to a portfolio of assets accords more closely
with individual and firm investment decision making. In order to deal with this revised focus, it is necessary to address the aggregation of the returns of the assets in the portfolio.

Consider a portfolio with only two assets whose portfolio shares are $w_1$ and $w_2$ respectively. The portfolio shares represent the fraction of the total portfolio value allocated to each of the assets with the normalisation condition

$$w_1 + w_2 = 1.$$ 

Using the definition of simple gross returns for each asset, the value of the portfolio between $t-1$ and $t$ may be calculated as

$$P_t = P_{t-1}w_1(1 + R_{1t}) + P_{t-1}w_2(1 + R_{2t}) = P_{t-1}(w_1(1 + R_{1t}) + w_2(1 + R_{2t})).$$

Rearranging slightly, this expression becomes

$$(1 + R_{Pt}) \equiv \frac{P_t}{P_{t-1}} = w_1(1 + R_{1t}) + w_2(1 + R_{2t}). \quad (1.13)$$

In words, the one-period gross return to a portfolio, $1 + R_{Pt}$, is given by the weighted sum of the gross returns to each of the assets using portfolio shares as weights. Expanding the right-hand side of equation (1.13) gives

$$1 + R_{Pt} = w_1 + w_1 R_{1t} + w_2 + w_2 R_{2t},$$

which yields the important result that for simple returns, the portfolio rate of return is equal to the weighted average of the returns to the assets

$$R_{Pt} = w_1 R_{1t} + w_2 R_{2t},$$

since $w_1 + w_2 = 1$. For $N$ assets the simple portfolio return is given by

$$R_{Pt} = \sum_{i=1}^{N} w_i R_{it}, \quad \sum_{i=1}^{N} w_i = 1. \quad (1.14)$$

This result does not extend to the case of log returns. From equation (1.5) and using the result in (1.14) it follows that the log return on a portfolio, $r_{Pt}$, is

$$r_{Pt} = \log(1 + R_{Pt}) = \log \left(1 + \sum_{i=1}^{N} w_i R_{it}\right) \neq \sum_{i=1}^{N} w_i r_{it}. \quad (1.15)$$

In most practical situations the fact that the log return to the portfolio is not the weighted sum of the log returns to the constituent assets is simply ignored. This is acceptable when the log returns are small, as is likely for short holding periods, in which case the log return on the portfolio is negligibly different to the weighted sum of the logarithm of the constituent asset returns because the approximation $r_{Pt} = \log(1 + R_{Pt}) \approx R_{Pt}$ is reasonably accurate when $R_{Pt}$ is small.
CHAPTER 1. FINANCIAL ASSET PRICES AND RETURNS

The result in equation (1.15) begs the question of how to combine log returns into the portfolio return. Consider again the case of two assets. Using the definition of log returns for each asset and expression (1.7), the value of the portfolio between \( t-1 \) and \( t \) may be calculated as

\[ P_t = P_{t-1} w_1 e^{r_{1t}} + P_{t-1} w_2 e^{r_{2t}}, \]

so that

\[ \log \left( \frac{P_t}{P_{t-1}} \right) \equiv r_{pt} = \log (w_1 e^{r_{1t}} + w_2 e^{r_{2t}}). \]

For \( N \) assets the log portfolio return is then

\[ r_{pt} = \log \left( \sum_{i=1}^{N} w_i e^{r_{it}} \right). \] (1.16)

More often than not, financial econometric work uses log returns and simply takes a weighted aggregate of these returns to find portfolio returns. This approach will also be used in Chapter 3 where simple portfolios are constructed using linear regression. The results above show that this commonly used aggregation procedure is not strictly correct.

Once returns, either simple returns or log returns, are available then equations (1.2) and (1.9) may be used for temporal aggregation of the portfolio returns. The situation is summarised in Table 1.4.

Table 1.4
Summary of expressions for computing portfolio returns using simple and log returns and how to aggregate portfolio returns to obtain the \( k \) period portfolio return.

<table>
<thead>
<tr>
<th>Aggregation</th>
<th>Simple Returns</th>
<th>Log Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portfolio Return</td>
<td>( R_{Pt} = \sum_{i=1}^{N} w_i R_{it} )</td>
<td>( r_{Pt} = \log \left( \sum_{i=1}^{N} w_i e^{r_{it}} \right) )</td>
</tr>
<tr>
<td>( K )-Period Return</td>
<td>( R_{Pt}(k) = \prod_{i=0}^{k-1} (1 + R_{Pt-i}) - 1 )</td>
<td>( r_{Pt}(k) = \sum_{i=0}^{k-1} r_{Pt-i} )</td>
</tr>
</tbody>
</table>

1.4 Stock Market Indices

A problem of particular importance is the return to a portfolio that comprises all or at least a selection of prominent stocks on a stock exchange. An aggregate summary measure of the performance of the stock market as a whole is known as a stock market index. Indices combine a selection of (a large number of) stocks in a particular way to create a portfolio. The index then represents the value of the portfolio and is expressed in terms of an average price
1.4. STOCK MARKET INDICES

that has been normalised in some way. Because stock market indices are price indices, the computation of returns to the index can be performed in exactly the same way as if it were a single stock.

The major stock market indices are constructed in one of two ways. Price-weighted indices construct a portfolio of all the stocks in the index in which one share of each of the stocks appears and the weight given to the share is therefore simply the price of the share. In other words, the total monetary value invested in each share is only proportional to the price of that share. Value-weighted indices construct a portfolio of all the stocks in the index in which the weight given to each stock is proportional to the total market value of its outstanding equity.

Figure 1.2: Daily observations on six international stock market indices for the period 4 January 1999 to 2 April 2014.
The six indices used most often in financial econometric work are plotted in Figure 1.2.

- Deutscher Aktien Index (DAX) comprises the 30 largest German companies that trade on the Frankfurt Stock Exchange. It is a value-weighted index although the weights are computed in a slightly more complex way than in a simple value weighting scheme.

- Dow Jones Industrial Average Index (Dow Jones or DJIA) is computed using 30 prominent United State corporations. The DJIA is a price-weighted index.

- Financial Times Stock Exchange 100 Index (FTSE) is a value-weighted index computed using the 100 largest companies listed on the London Stock Exchange.

- Hang Seng Index (Hang Seng or HSX) comprises 40 of the largest companies that trade on the Hong Kong Exchange. It is a value-weighted index.

- Nikkei 225 Index (Nikkei or NKX) is a price-weighted index made up of 225 prominent companies listed on the Tokyo Stock Exchange.

- Standard and Poors Composite 500 (S&P 500) is a market-value weighted index. The index is computed by summing the market value of the outstanding equity in each firm in the index.

The falls in these market indices that occurred around the collapse of the dot-com bubble in the early 2000s and the global financial crisis of 2008-2009 are evident in each of the graphs plotted in Figure 1.2.

In addition to these six indices, another commonly encountered index is the NASDAQ Composite Index, which is a value-weighted index of all the stocks listed on the NASDAQ stock exchange. It is usually regarded as an index of the performance of technology companies and is particularly associated with the dot-com bubble of the late 1990s, which created and destroyed some $8 trillion dollars of shareholder wealth over a period of 5−6 years.

Table 1.5 lists the 30 component stocks of the Dow Jones Index obtained from Bloomberg in September 2014. The monthly closing price for December 2013 is also listed together with the market capitalisation (US$ bill.) of the component stocks (price of share × number of outstanding shares). Despite the fact that the DJIA is a price-weighted index, Table 1.5 also shows the notional share that each stock would have in a value-weighted index.

The DJIA is computed as

\[ DJIA_t = \frac{1}{D} \sum_{j=1}^{30} P_{jt}, \]
where the quantity $D$ which appears in the denominator is known as the Dow Jones divisor. The divisor started out as the number of stocks in the index so the DJIA was a simple average, but subsequent adjustment due to stock splits and structural changes required the divisor to be adjusted in order to preserve the continuity of the index. For example, the appropriate value of the divisor in December 2013 was 0.15571590501117.

Using the closing prices in Table 1.5, the DJIA for December 2013 is computed as

$$
\text{DJIA}_{\text{Dec13}} = \frac{140.25 + 90.73 + 35.15 + \cdots + 222.68 + 78.69 + 76.40}{0.15571590501117} = \frac{2581.25}{0.15571590501117} = 16576.662,
$$

which is identical to the value of the index, 16576.66, quoted by Bloomberg for December 2013. The DJIA is a price-weighted average. The main advantage of price weighting is its simplicity but its primary disadvantage is that stocks with the highest prices, like Visa ($222.68), IBM ($187.57) and Goldman Sachs ($177.26), have a greater relative impact on the index than perhaps they should have.

The other major type of weighting scheme is to weight stocks by market capitalisation, giving a value weighted average. In consequence, stocks like Exxon (0.094), Microsoft (0.066) and General Electric (0.060), would have the largest weights in the index if they were value-weighted. The primary disadvantage of value weighting is that constituent securities whose prices have risen the most (or fallen the most) have a greater (or lower) weight in the index. This weighting method can potentially lead to overweighting stocks that have risen in price (and may be overvalued) and underweighting stocks that have declined in price (and may be undervalued).

The differences between price weighting and value weighting are illustrated in Figure 1.3 in which the 30 constituent stocks of the Dow Jones are combined to form two hypothetical indices, one based on simple price weighting and the other using shares constructed from market capitalisation as shown in Table 1.5. Both indices in Figure 1.3 are normalised to take the value 100 in January 1990. While the price-weighted and value-weighted indices track each other fairly closely over the period, the price-weighted index seems to over-emphasise market movements during the period of the dot-com bubble during the latter half of the 1990s as well as the speed of the recovery from the 2008 global financial crisis.
Table 1.5

The 30 United States stocks used in the construction of the Dow Jones Index. Month-end closing prices adjusted for splits and dividends and quoted in United States $ are shown for the month of December 2013 together with total outstanding value of the company’s shares ($Bill.).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3M Co.</td>
<td>MMM</td>
<td>140.250</td>
<td>93.300</td>
<td>0.020</td>
</tr>
<tr>
<td>American Express Co.</td>
<td>AXP</td>
<td>90.730</td>
<td>97.196</td>
<td>0.021</td>
</tr>
<tr>
<td>AT&amp;T Inc.</td>
<td>T</td>
<td>35.150</td>
<td>97.196</td>
<td>0.021</td>
</tr>
<tr>
<td>The Boeing Co.</td>
<td>BA</td>
<td>136.490</td>
<td>102.566</td>
<td>0.022</td>
</tr>
<tr>
<td>Caterpillar Inc.</td>
<td>CAT</td>
<td>90.810</td>
<td>57.787</td>
<td>0.012</td>
</tr>
<tr>
<td>Chevron Corp.</td>
<td>CVX</td>
<td>124.910</td>
<td>240.224</td>
<td>0.051</td>
</tr>
<tr>
<td>Cisco Systems Inc.</td>
<td>CSCO</td>
<td>22.450</td>
<td>120.032</td>
<td>0.025</td>
</tr>
<tr>
<td>The Coca-Cola Co.</td>
<td>KO</td>
<td>41.310</td>
<td>182.422</td>
<td>0.039</td>
</tr>
<tr>
<td>El du Pont de Nemours &amp; Co.</td>
<td>DD</td>
<td>64.970</td>
<td>60.169</td>
<td>0.013</td>
</tr>
<tr>
<td>Exxon Mobil Corp.</td>
<td>XOM</td>
<td>101.200</td>
<td>442.094</td>
<td>0.094</td>
</tr>
<tr>
<td>General Electric Co.</td>
<td>GE</td>
<td>28.030</td>
<td>283.590</td>
<td>0.060</td>
</tr>
<tr>
<td>The Goldman Sachs Group Inc.</td>
<td>GS</td>
<td>177.260</td>
<td>83.353</td>
<td>0.018</td>
</tr>
<tr>
<td>The Home Depot Inc.</td>
<td>HD</td>
<td>82.340</td>
<td>115.953</td>
<td>0.025</td>
</tr>
<tr>
<td>Intel Corp.</td>
<td>INTC</td>
<td>25.960</td>
<td>129.047</td>
<td>0.027</td>
</tr>
<tr>
<td>International Business Machine</td>
<td>IBM</td>
<td>187.570</td>
<td>203.674</td>
<td>0.043</td>
</tr>
<tr>
<td>Johnson &amp; Johnson</td>
<td>JNJ</td>
<td>91.590</td>
<td>258.415</td>
<td>0.055</td>
</tr>
<tr>
<td>JPMorgan Chase &amp; Co.</td>
<td>JPM</td>
<td>58.480</td>
<td>219.837</td>
<td>0.047</td>
</tr>
<tr>
<td>McDonalds Corp.</td>
<td>MCD</td>
<td>97.030</td>
<td>96.548</td>
<td>0.020</td>
</tr>
<tr>
<td>Merck &amp; Co. Inc.</td>
<td>MRK</td>
<td>50.050</td>
<td>146.242</td>
<td>0.031</td>
</tr>
<tr>
<td>Microsoft Corp.</td>
<td>MSFT</td>
<td>37.430</td>
<td>312.464</td>
<td>0.066</td>
</tr>
<tr>
<td>NIKE Inc.</td>
<td>NKE</td>
<td>78.640</td>
<td>69.955</td>
<td>0.015</td>
</tr>
<tr>
<td>Pfizer Inc.</td>
<td>PFE</td>
<td>30.630</td>
<td>198.515</td>
<td>0.042</td>
</tr>
<tr>
<td>The Procter &amp; Gamble Co.</td>
<td>PG</td>
<td>81.410</td>
<td>221.291</td>
<td>0.047</td>
</tr>
<tr>
<td>The Travelers Companies Inc.</td>
<td>TRV</td>
<td>90.540</td>
<td>32.963</td>
<td>0.007</td>
</tr>
<tr>
<td>United Technologies Corp.</td>
<td>UTX</td>
<td>113.800</td>
<td>104.421</td>
<td>0.022</td>
</tr>
<tr>
<td>United Health Group Inc.</td>
<td>UNH</td>
<td>75.300</td>
<td>75.809</td>
<td>0.016</td>
</tr>
<tr>
<td>Verizon Communications Inc.</td>
<td>VZ</td>
<td>49.140</td>
<td>140.626</td>
<td>0.030</td>
</tr>
<tr>
<td>Visa Inc.</td>
<td>V</td>
<td>222.680</td>
<td>141.756</td>
<td>0.030</td>
</tr>
<tr>
<td>Wal-Mart Stores Inc.</td>
<td>WMT</td>
<td>78.690</td>
<td>254.623</td>
<td>0.054</td>
</tr>
<tr>
<td>The Walt Disney Co.</td>
<td>DIS</td>
<td>76.400</td>
<td>134.256</td>
<td>0.028</td>
</tr>
</tbody>
</table>

Source: Bloomberg.
Figure 1.3: The effect of price weighting and value weighting on an index comprising 30 stocks that make up the Dow Jones Industrial Average. The indices are computed using monthly data on prices and market capitalisation for the period January 1990 to December 2013 with each index scaled to start from 100.

1.5 Bond Yields

As noted in Section 1.2, zero-coupon bonds may be created from coupon paying bonds by separating the coupons from the principal and trading each of these components independently in the process known as stripping. Consequently, much of the econometric analysis of the bond market uses data based on zero-coupon bonds. The critical concept when dealing with bonds, which relates to the return on a stock, is the yield to maturity. If a zero coupon bond has a face value of $1 paid at maturity, $n$, the price of the bond purchased at time $t$ equals the discounted present value of the principal, which is given by

$$P_{nt} = 1 \times \exp \left(-ny_{nt}\right),$$  \hspace{1cm} (1.17)

where $y_{nt}$ is the discount rate or yield and is commonly expressed in per annum terms. The yield on a bond is therefore the discount rate that equates the present value of the bond’s face value to its price.

Taking natural logarithms and rearranging equation (1.17) gives

$$y_{nt} = -\frac{1}{n} \log P_{nt} = -\frac{1}{n} p_{nt}.$$  \hspace{1cm} (1.18)
This expression shows that the yield is inversely proportional to the natural logarithm of the price of the bond, where the proportionality constant is $-1/n$. Moreover, as the price of the bond $P_{nt}$ is always less than $1$ then from the properties of logarithms, $p_{nt}$ is a negative number and the yield in equation (1.18) will always be positive.

Figure 1.4: Scatter plots of observed yields for the months of March, May, July and August 1989 for United States zero coupon bonds.

Governments issue bonds of differing lengths to maturity. Bonds at the shorter end of the maturity spectrum (maturity less than 12 months) are generally zero-coupon bonds (bonds that pay no coupon or interest), while the coupon bonds (for which the holder receives regular interest payments) can have a maturity as long as 30 years. The term structure of interest rates is the relationship between time to maturity and yield to maturity and the yield curve is a plot of the term structure of yield to maturity against time to maturity at a specific time. Figure 1.4 presents scatter plots of observed United States zero-coupon bond yield curves for the months of March, May, July and August 1989, for yields ranging from 1 to 120 months. The yields are computed from the end-of-month price quotes taken from the CRSP government bonds files, the same data as that used in Diebold and Li (2006).
The plots of the yield curves in Figure 1.4 reveal a few well-known features.

1. At any point in time when the yield curve is observed, all the maturities may not be represented. This is particularly true at longer maturities where the number of observed yields is much sparser than at the short end of the maturity spectrum.

2. The yields at longer maturities tend to be less volatile than the yields at the shorter end of the maturity spectrum.

3. On the assumption that longer-term financing should carry a risk premium, a natural expectation would be for the yield curve to slope upward. However, the empirical plots in Figure 1.4 show that the yield curve can in practice assume a variety of shapes, including upward sloping, downward sloping, humped and even inverted humped. These shapes are ultimately determined by the demand and supply of bonds of various maturities, market expectations, and risk assessments.

Modelling bond yields and the term structure are important and often challenging tasks in financial econometric work. Various aspects of these tasks in modelling bond yields and the tools designed to address them are examined in Chapters 6, 9 and 12.

1.6 Exercises

The data required for the exercises are available for download as EViews work-files (*.wf1), Stata datafiles (*.dta), comma delimited text files (*.csv) and as Excel spreadsheets (*.xlsx).

1. **Equity Prices, Dividends and Returns**

   (a) Log on to Yahoo Finance (https://au.finance.yahoo.com) and load the current quoted prices for The Boeing Company (BA). Compare the current situation with that reported in Table 1.1.

   (b) Observe the historical daily prices for Boeing. What do you notice about the order in which they are presented?

   (c) Examine the daily prices and scrutinise the days on which dividend payments are made. Verify that the dividend adjustments made to the historical price series are correct.

   (d) Obtain monthly price data on Boeing. Are the quoted monthly prices beginning or end of month quotes?

2. **Simple and Logarithmic Returns**
CHAPTER 1. FINANCIAL ASSET PRICES AND RETURNS

capm.*

The data are monthly observations for the period April 1990 to July 2004 on the equity prices of Exxon, General Electric, IBM and Microsoft and Walmart, together with the price of Gold, the S&P 500 index and a short-term interest rate.

(a) Plot the price indices and comment on the results.

(b) Compute simple and logarithmic returns to each of the assets. For each asset plot the two returns series and comment on any differences.

(c) Compute the simple and logarithmic returns to each of the assets over the entire sample period and comment on the difference.

(d) Assume that you hold each of the stocks in a portfolio. Compute the portfolio returns in both simple and logarithmic form for the first seven months of 2004.

3. Returns

Djindexstocks.*

The data are monthly observations for the period 31 January 1990 to 31 December 2013 on the Dow Jones Index, the prices and market capitalisation of the 30 constituent stocks of the index, and the risk-free interest rate.

(a) Consider the historical prices for Microsoft for the years 2012 and 2013. For these two years, compute the price relative, simple and logarithmic monthly returns, and simple and logarithmic annualised returns. Compare your results with Table 1.3.

(b) Compute the logarithmic and simple returns to holding each of the 30 stocks in the Dow Jones for the month of December 2012.

(c) Assuming equal shares, compute the simple and logarithmic returns to holding a portfolio comprising each of the 30 Dow Jones stocks for the month of December 2012.

4. Stock Indices
1.6. EXERCISES

The data are daily observations on the Dow Jones, S&P 500, Hang Seng, Nikkei, Dax and FTSE stock indices for the period 4 January 1999 to 2 April 2014.

(a) Plot the indices. Compare your results with Figure 1.2.
(b) Compute the daily logarithmic and simple returns of each of the indices and plot them. Comment on any differences.
(c) Express the daily logarithmic and simple returns in annualised form and plot the resultant series. Comment on your results.
(d) Compute the returns to holding each of the indices over the entire sample period in both logarithmic and simple form. Comment on the results.

5. Dow Jones Index

The data are monthly observations for the period 31 January 1990 to 31 December 2013 on the Dow Jones Index, the prices and market capitalisation of the 30 constituent stocks of the index, and the risk-free interest rate.

(a) Compute the Dow Jones Industrial Average for December 2013 using

\[
DJIA_t = \frac{1}{D} \sum_{j=1}^{30} P_{jt}.
\]

where the Dow Jones divisor, \(D\), is taken to be 0.15571590501117. Verify that your result is identical to quoted value of the Dow for that month.

(b) Construct portfolio shares for each of the Dow Jones stocks based on market capitalisation for the month of December 2013. Comment on which stocks that receive the most weight in the Dow under the price and market capitalisation weighting schemes, respectively.

(c) Combine the 30 constituent stocks of the Dow Jones to form two indices, one based on simple price weighting and the other using shares constructed from market capitalisation. Plot the indices over the sample period and comment on the differences.

6. Australian Stocks
The data are monthly observations on the prices of the largest 136 stocks in Australia from December 1999 to June 2014. Consider a portfolio constructed by holding one share in every one of the \( N = 136 \) stocks in the dataset that records a price, \( P_{jt} \) at every time \( t \) in the sample period.

(a) Compute the simple and log returns to the portfolio over the full sample period using

\[
R(P) = \frac{P_T}{P_1} - 1, \quad r(P) = \log \left( \frac{P_T}{P_1} \right),
\]

respectively, in which

\[
P_t = \sum_{i=1}^{N} P_{it}
\]

Comment on the results.

(b) Compute the portfolio weights of each stock in the portfolio for every time \( t \) using the formula

\[
w_{it} = \frac{P_{it}}{\sum_{i=1}^{N} P_{it}},
\]

in which \( N \) is the number of stocks in the portfolio.

(c) Compute the simple return and log returns to the portfolio in each time period, respectively,

\[
R_{Pt} = \sum_{i=1}^{N} w_{it-1} R_{it}, \quad r_{Pt} = \log \left( \sum_{i=1}^{N} w_{it-1} e^{r_{it}} \right),
\]

remembering to use the weight at the beginning of the holding period.

(d) Compare the results obtained in (a) and (c).
Chapter 2

Properties of Financial Data

The financial pages of newspapers and magazines, online financial sites, and academic journals all routinely report a plethora of financial statistics. Even within a specific financial market, the data may be recorded at multiple observation frequencies and the same data may be presented in various ways. As will be seen, the time series based on these representations have very different statistical properties and reveal distinct features of the underlying phenomena relating to both long run and short run behaviour.

The characteristics of financial data may also differ across markets. For example, there is no reason to expect that equity markets behave the same way as currency markets, or for commodity markets to behave the same way as bond markets. In some cases, like currency markets, trading is a nearly continuous activity, while other markets open and close in a regulated manner according to specific times and days. Real estate markets differ further in terms of their intermittent transactions, limited liquidity, and supply side constraints. Yet their importance to other financial markets such as the mortgage market and wider linkages to the macroeconomy are now especially evident in the aftermath of the 2008 financial crisis. Options markets also have their own special characteristics and offer a wide and growing range of financial instruments that relate to other financial assets and markets.

One important preliminary role of statistical analysis is to find stylised facts that characterise different types of financial data and particular markets. Such analysis is primarily descriptive and helps us to understand the prominent features of the data and the differences that can arise from basic elements like varying the sampling frequency and implementing various transformations. Accordingly, the primary aim of this chapter is to highlight the main characteristics of financial data and establish a set of stylised facts for financial time series. These characteristics will be used throughout the book as important inputs in the building and testing of financial models.
2.1 A First Look at the Data

This section identifies the key empirical characteristics of financial data. Special attention is devoted to establishing a set of stylised empirical facts that characterise financial data. These empirical characteristics are important for building financial models and for developing appropriate methods of inference.

2.1.1 Prices

Figure 2.1 gives a plot of the monthly United States equity price index (S&P 500) for the period January 1950 to September 2016. The time path of equity prices shows long-run growth over this period whose general shape is captured by the exponential trend graphed against the data in the figure. While the exponential trend prescribes a general pattern of movement in prices over this historical period, it is also clear that there are major swings in which prices wander above and below this exponential trend for sustained periods. This observed exponential pattern in the equity price index may be expressed formally as

\[ P_t = P_{t-1} \exp(r_t), \]  

(2.1)

where \( P_t \) is the current equity price, \( P_{t-1} \) is the previous month’s price and \( r_t \) is the rate of the increase between month \( t-1 \) and month \( t \).

If \( r_t \) in (2.1) is restricted to take the same constant value, \( r \), in all time periods, then equation (2.1) becomes

\[ P_t = P_{t-1} \exp(r). \]  

(2.2)

The relationship between the current price, \( P_t \) and the price two months earlier, \( P_{t-2} \), is

\[ P_t = P_{t-1} \exp(r) = P_{t-2} \exp(r) \exp(r) = P_{t-2} \exp(2r). \]

By continuing this recursion, the relationship between the current price, \( P_t \), and the price \( t \) months earlier, \( P_0 \), is given by

\[ P_t = P_0 \exp(rt). \]  

(2.3)

It is this exponential function that is plotted in Figure 2.1 in which \( P_0 = 16.88 \) is the equity price in January 1950 and the constant growth rate in monthly prices is set at \( r = 0.006 \).

The exponential function in equation (2.3) provides a within-sample predictive relationship based on the observed long-run growth behaviour over 1950-2016. If an investor in January 1950 assumed that the long-run exponential growth rate was 0.006, then this investor could predict the price of
2.1. A FIRST LOOK AT THE DATA

Figure 2.1: Monthly S&P 500 equity price index for the United States from January 1950 to September 2016. Fitted values (dashed line) are obtained using the exponential model in equation (2.3) with $r = 0.006$.

equities in September 2016. As there are 800 months between these dates, the forecast price for September 2016 is

$$P(\text{Sep.2016}) = 16.88 \times \exp(0.006 \times 800) = 2051.10.$$  

The actual equity price in September 2016 for the data plotted in Figure 2.1 is 2157.69 showing that the predicted price is fairly accurate with a percentage forecast error of

$$100 \times \frac{2157.69 - 2051.10}{2157.69} = 4.94\%.$$  

Forecasting methods for financial variables are discussed in Chapter 7.

An alternative way of analysing the long run time series behaviour of asset prices is to plot the logarithm of prices over time. An example is given in Figure 2.2 where the natural logarithm of the equity price given in Figure 2.1 is presented. Comparing the two series shows that, while prices increase at an increasing rate (Figure 2.1), the logarithm of price increases at a constant rate (Figure 2.2). To see why this is the case, we take natural logarithms of equation (2.3) to yield

$$p_t = p_0 + rt, \quad (2.4)$$

where lower case letters now denote the natural logarithms of the variables, namely, $p_t = \log P_t$ and $p_0 = \log P_0$. This is a linear equation between $p_t$ and $t$ in which the slope is equal to the constant $r$. This equation also forms the
basis of the definition of log returns, a point that is now developed in more detail.

![Log price chart](image)

**Figure 2.2:** The natural logarithm of the monthly equity price index for the United States from January 1950 to September 2016. The dashed line is a linear function with slope 0.006.

### 2.1.2 Returns

Figure 2.3 plots monthly log equity returns for the United States defined as
\[ r_t = \log(P_t) - \log(P_{t-1}) = p_t - p_{t-1}, \]
over the period January 1950 to September 2016 (see Chapter 1 for more details on computing returns). The log returns are seen to hover around a return value that is near zero over the sample period. This value is in fact \( r = 0.006 \), which is the estimate used in the earlier computations in Section 2.1.1. This feature of equity returns contrasts dramatically with the trending character of the corresponding equity prices presented in Figure 2.1.

The marked empirical difference in the two series for prices and returns reveals an interesting aspect of stock market behaviour. It is often emphasised in the financial literature that investment in equities should be based on long run considerations rather than the prospect of short run gains. The reason is that stock prices can be very volatile in the short run. This short run behaviour is reflected in the high variability of the stock returns shown in Figure 2.3. Yet, although stock returns hover around a value of approximately
zero, stock prices (which accumulate these returns) tend to trend noticeably upwards over time, as is apparent in Figure 2.1. This tendency of stock prices to drift upwards over time is taken up again in Chapter 5. For present purposes, it is sufficient to remark that when returns are measured over very short periods of time, any tendency of prices to drift upwards is virtually imperceptible because that effect is so small and is swamped by the apparent volatility of the returns. This interpretation puts emphasis on the fact that returns generally focus on short run effects whereas price movements can trend noticeably upwards (or downwards) over extended periods of time.

### 2.1.3 Dividends

In many applications in finance, as in economics, the focus is on understanding the relationships among two or more series. For instance, in present value models of equities, the price of an equity is equal to the discounted future stream of dividend payments

\[
P_t = E_t \left[ \frac{D_{t+1}}{(1 + \delta_t)} + \frac{D_{t+2}}{(1 + \delta_t)^2} + \frac{D_{t+3}}{(1 + \delta_t)^3} + \cdots \right],
\]

where \( E_t(D_{t+n}) \) represents the expectation of dividends in the future at time \( t + n \) given information available at time \( t \) and \( \delta_t \) is the discount rate at time \( t \).
The relationship between equity prices and dividends is highlighted in Figure 2.4 which plots United States equity prices and dividend payments from January 1950 to September 2016. There appears to be a relationship between the two series as both series exhibit positive exponential-like trends with intermittent downturns that are more striking for equity prices.

To analyse the relationship between equity prices and dividends more closely, assume that current expectations of future dividends are given by present dividends, $E_t(D_{t+n}) = D_t$. The present value relationship in equation (2.5) can now be written as

$$P_t = D_t \left( \frac{1}{1 + \delta_t} + \frac{1}{(1 + \delta_t)^2} + \ldots \right)$$

$$= \frac{D_t}{1 + \delta_t} \left( 1 + \frac{1}{1 + \delta_t} + \frac{1}{(1 + \delta_t)^2} + \ldots \right)$$

$$= \frac{D_t}{1 + \delta_t} \left( \frac{1}{1 - 1/(1 + \delta_t)} \right)$$

$$= \frac{D_t}{\delta_t}, \quad (2.6)$$
A FIRST LOOK AT THE DATA

where the penultimate step uses the sum of a geometric progression. Rearranging this expression gives

\[ \delta_t = \frac{D_t}{P_t}, \quad (2.7) \]

which defines the dividend yield plotted in Figure 2.5 based on the data in Figure 2.4. The dividend yield exhibits no upward trend and instead wanders randomly around the level of 0.03. This behaviour is in stark contrast to the equity price and dividend series which both exhibit a strong tendency to trend upwards.

Figure 2.5: Monthly United States dividend yield for the period December 1950 to September 2016.

The example of the dividend yield illustrates how combining two or more series in an appropriate way can change the time series properties of the data - in the present case by apparently eliminating the strong upward trending behaviour in two series by standardising one series in terms of the other. The process of combining trending financial variables into new variables that do not exhibit trends is a form of trend reduction by means of variable combination. This idea is explored in more detail in Chapter 6.

An alternative representation of the present value model suggested by equa-

---

1Recall that an infinite geometric progression is summed as

\[ 1 + \lambda + \lambda^2 + \lambda^3 + \ldots = \frac{1}{1-\lambda}, \quad |\lambda| < 1, \]

where in the example \( \lambda = 1/(1+\delta_t) \).
tion (2.6) is to take natural logarithms and rearrange for log \( P_t \) to give

\[ \log P_t = -\log \delta_t + \log D_t. \]

Assuming equities are priced according to the present value model, this equation reveals a linear relationship between \( \log P_t \) and \( \log D_t \).

### 2.1.4 Bond Yields

Figure 2.6 gives plots of yields on United States zero coupon bonds for maturities of 3 and 9 months. The yields to the different maturities are not distinguished from one another as the primary purpose is to discover some of the time series properties of bond yields.

![Figure 2.6: Monthly United States zero coupon bond yields for maturities of 3 months (solid line) and 9 months (dashed line) over the period December 1946 to February 1987. The different maturities are hardly distinguishable, emphasising the fact that yields of different maturities track each other closely.](image)

The bond yield plots shown in Figure 2.6 reveal three important properties.

1. The yields are increasing over time, so they exhibit some form of trending behaviour. This feature of financial time series is the subject matter of Chapter 5.
2. The variance of the yields tends to grow as the levels of the yields increase. This is called the levels effect and is investigated in more detail in Chapter 9.

3. The yields of different maturities follow one another closely, indeed so closely that they are hardly distinguishable in the scale of Figure 2.6. This is yet another example of variables that exhibit trending behaviour and move commonly over time.

![Spread between the United States 9-month and 3-month bond yields for the period December 1946 to February 1987.](image)

Property 3 of bond yields may be highlighted by computing the spread between the yields on a long maturity bond and a short maturity bond. Figure 2.7 gives the spread between the long maturity bond (9 months) and the short maturity bond (3 months). Comparison of Figures 2.6 and 2.7 reveals that yields exhibit vastly different time series patterns to spreads, with the latter showing no evidence of trends. However, there is still evidence that the variability of the spread is not constant over the sample, a feature that may be interpreted as a form of trend behaviour in the variance.

### 2.1.5 Financial Distributions

An important but limiting assumption underlying many theoretical models and empirical methods in finance is that returns are normally distributed.
This assumption has been widely used in portfolio allocation models, in pricing options, and in many other applications. An example of an empirical returns distribution is given in Figure 2.8, which gives the histogram of hourly United States exchange rate returns computed relative to the British pound. Even though this distribution exhibits some characteristics that are consistent with a normal distribution such as symmetry, the distribution differs from normality in two important ways, namely, the presence of a sharp peak in the centre of the distribution and evidence of slightly heavier tails. This feature suggests that there are many more observations where the exchange rate hardly moves and for which there are a greater number of smaller returns than there would be if returns were drawn randomly from a normal population. Distributions exhibiting these properties are called leptokurtic.

Figure 2.8: Empirical distribution of hourly $/£ exchange rate returns for the period 1 January 1986 00:00 to 15 July 1986 11:00 with a normal distribution overlaid.

The example given in Figure 2.8 is for exchange rate returns. But the property of heavy tails and peakedness of the distribution of returns is common for other asset markets including equity and commodity markets. All of these empirical distributions are therefore inconsistent with the assumption of normality in terms of these empirical characteristics. Financial models that are derived from normal distributions may therefore result in financial instruments such as options being incorrectly priced or measures of risk being underestimated due to their failure to incorporate these features of asset returns.
2.1.6 Transactions

A property of all of the financial data analysed so far is that observations on a particular variable are recorded at discrete and regularly spaced points in time. The data on equity prices and dividend payments in Figure 2.4 and the data on zero coupon bond yields in Figure 2.6 are all recorded every month. In fact, higher frequency data are also available at daily and hourly frequencies.

More recently, high-frequency transactions data which records the price of every trade conducted during the trading day has become available. An example is given in Table 2.1 which gives a snapshot of the trades recorded for American Airlines on 1 August 2006 at 9:42 am. The variable Trade, \( x_t \), is a binary variable signifying whether a trade has taken place at time \( t \) so that

\[
x_t = \begin{cases} 
1 & : \text{Trade occurs} \\
0 & : \text{No trade occurs.}
\end{cases}
\]

<table>
<thead>
<tr>
<th>Sec.</th>
<th>Trade</th>
<th>Duration</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>$21.58</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>.</td>
<td>$21.58</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>.</td>
<td>$21.58</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>.</td>
<td>$21.58</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>.</td>
<td>$21.58</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>.</td>
<td>$21.58</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>6</td>
<td>$21.59</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>1</td>
<td>$21.59</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>.</td>
<td>$21.59</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>2</td>
<td>$21.59</td>
</tr>
</tbody>
</table>

The duration between trades is measured in seconds, and the corresponding price of the asset at the time of the trade is also recorded. Table 2.1 shows that there is a trade at the 5 second mark where the price is $21.58. The next trade occurs at the 11 second mark at a price of $21.59, so the duration between trades is 6 seconds. There is another trade straight away at the 12 second mark at the same price of $21.59, in which case the duration is just 1 second. There is no trade in the following second, but there is another two seconds later at the 14 second mark, again at the same price of $21.59, so the duration is 2 seconds.
The time differences between trades of American Airlines (AMR) shares is further highlighted by the histogram of the duration times given in Figure 2.9. This distribution has an exponential shape with the duration time of 1 second, being the most common. However, there are a number of durations in excess of 25 seconds and there are some times even in excess of 50 seconds.

![Empirical distribution of durations (in seconds) between trades of American Airlines (AMR) on 1 August 2006 from 9:30 am to 4:00 pm.](image)

The important feature of transactions data that distinguishes it from the time series data discussed earlier is that the time interval between trades is not regular or equally spaced. In fact, if high frequency data are used, such as 1 minute data, there will be periods where no trades occur in the window of time and the price will not change. This is especially so in thinly traded markets. The implication of using such transactions data is that the models specified in econometric work need to incorporate those features, including the apparent randomness in the observation interval between trades. Such features end up being important in applied work because they connect most closely with the institutional structure of the market and the manner in which trades are conducted. Correspondingly, the appropriate statistical techniques for analysing such data are expected to be different from the techniques used to analyse regularly spaced financial time series data. These issues apply for very high frequency, irregularly spaced data and are investigated further in Chapter 16 on financial microstructure effects.
2.2 Summary Statistics

In the previous section, the time series properties of financial data are explored using graphical tools of line charts and histograms in conjunction with tabulation. This section introduces some simple descriptive statistical methods to summarise the main characteristics of the data. While these methods are useful summary measures in general, there are situations where they are inappropriate and a few examples are given to highlight such cases where these simple measures may break down or fail to reveal important characteristics of the data.

2.2.1 Univariate

In assessing investment opportunities, it is natural for investors to seek summary information about the financial assets of interest to them. In particular, investors would like to know about the return that is to expected from investing in the asset, as well as the risk of the investment where risk refers to the uncertainty surrounding the value or payoff from investing. In addition, the information about the relatively likelihood of extreme returns is also of interest. This information is provided in terms of the following summary statistics.

Sample Mean

A simple measure of the expected return is given by the sample mean

\[ \bar{r} = \frac{1}{T} \sum_{t=1}^{T} r_t. \]

The returns to monthly S&P 500 data, \( r_t \), are plotted for the period January 1950 to September 2016 in Figure 2.3. The sample mean of these data is \( \bar{r} = 0.006 \) per month. Expressed in annual terms, the mean return is \( 0.006 \times 12 = 0.072 \) so that the average return over the period 1950 to 2016 is 7.2% per annum. The sample mean represents the level around which \( r_t \) fluctuates and therefore represents a summary measure of the location of the data.

An example where the sample mean is an inappropriate summary measure occurs when the data are trending. Figure 2.10 plots the equity price index with the sample mean of prices, \( \bar{P} = 501.5 \), superimposed. The sample mean, \( \bar{P} \), no longer represents a useful long-run level about which \( P_t \) is located and is therefore not an suitable summary measure of the location of the data. In fact, the price data spend very little time around this sample mean level, as is evident in Figure 2.10.
Sample Variance and Standard Deviation

A measure of the deviation of the actual return on an asset around its sample mean is given by the sample variance.

\[ s^2 = \frac{1}{T} \sum_{t=1}^{T} (r_t - \bar{r})^2. \]

This form of the sample variance is a biased estimator of the population variance. An unbiased estimator is to replace the \( T \) in the denominator with \( T - 1 \) which is known as a degrees of freedom or small sample correction. In most financial econometric applications, the sample size \( T \) is large enough for this difference to be negligible. In the case of the S&P 500 returns data, the sample variance is \( s^2 = 0.0348^2 = 0.0012 \).

In finance, the sample standard deviation\(^2\), which is the square root of the sample variance,

\[ s = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (r_t - \bar{r})^2}, \]

\(^2\text{The terminology standard deviation was introduced by a famous English scientist and polymath, Sir Francis Galton (1822-1911), who conceived this measure as a quantity to represent dispersion in the same measurement scale as the data.}\)
is often used as a measure of the riskiness of an investment and is called the volatility of a financial return. For the S&P 500 data the standard deviation or volatility of returns is \( s = 0.0348 \). In the present context, the standard deviation of returns is therefore easily interpretable in terms of the returns data, whereas the sample variance has the scale of returns squared.

**Sample Skewness**

If the extreme returns in any sample are mainly positive (negative), the distribution of \( r_t \) is positively (negatively) skewed. A measure of skewness in the sample is

\[
SK = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{r_t - \bar{r}}{s} \right)^3.
\]

If the sample skewness is zero, then the distribution is said to be symmetric. Figure 2.11 gives a histogram of the United States equity returns previously plotted in Figure 2.3. The sample skewness is computed to be \( SK = -1.005 \), where the sign of the statistic emphasises negative skewness. The result is supported by the evidence of a heavier left tail in the distribution of returns in comparison with a normal distribution, as shown in Figure 2.11.

![Figure 2.11: Empirical distribution of monthly United States equity returns for the period January 1950 to September 2016. Superimposed on the returns distribution is the best fitting normal distribution.](image-url)
Sample Kurtosis

If there are extreme returns relative to a benchmark distribution (usually the normal distribution), the distribution of \( r_t \) is said to exhibit excess kurtosis, which is named after a Greek word \textit{kurtos} meaning ‘arching or bulging’. A measure of kurtosis in the sample is

\[
KT = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{r_t - \bar{r}}{s} \right)^4.
\]

Comparing this value to \( KT = 3 \), which is the kurtosis value of a normal distribution, gives a measure of excess kurtosis

\[
\text{EXCESS } KT = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{r_t - \bar{r}}{s} \right)^4 - 3.
\]

In the case of the United States log equity returns, the sample kurtosis is \( KT = 6.938 \). This value is much greater than 3, revealing that there are more extreme returns in the data (or heavier tails) than would be predicted by the normal distribution.

2.2.2 Bivariate

The statistical measures discussed so far summarise the characteristics of the returns to a single asset. Perhaps even more important in finance is understanding the interrelationships between two or more financial assets. For example, in constructing a diversified portfolio, the aim is to include assets whose fluctuations in returns do not match each other perfectly. In this way, the value of the portfolio is partially protected even though there will be certain assets in the portfolio that are performing poorly.

Covariance

A measure of co-movements between the returns on two assets, \( r_{it} \) and \( r_{jt} \), is the sample covariance given by

\[
s_{ij} = \frac{1}{T} \sum_{t=1}^{T} (r_{it} - \bar{r}_i) (r_{jt} - \bar{r}_j),
\]

in which \( \bar{r}_i \) and \( \bar{r}_j \) are the respective sample means of the returns on assets \( i \) and \( j \). A positive covariance, \( s_{ij} > 0 \), shows that the returns of asset \( i \) and asset \( j \) have a tendency to move together. That is, when the return on asset \( i \) is above its mean, the return on asset \( j \) is also likely to be above its mean. A negative covariance, \( s_{ij} < 0 \), indicates that when the returns of asset \( i \) are above
its sample mean, on average, the returns on asset \( j \) are likely to be below its sample mean. Covariance has a particularly important role to play in empirical finance, as will become clear in Chapter 3.

**Correlation**

Another measure of association is the correlation coefficient \(^3\) given by

\[
\rho_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}},
\]

in which

\[
s_{ij} = \frac{1}{T} \sum_{t=1}^{T} (r_{it} - \bar{r}_i)^2,
\]

\[
s_{jj} = \frac{1}{T} \sum_{t=1}^{T} (r_{jt} - \bar{r}_j)^2,
\]

represent the respective variances of the returns of assets \( i \) and \( j \). The correlation coefficient is the covariance scaled by the standard deviations of the two returns. The correlation has the property that it has the same sign as the covariance, as well as the additional property that it lies in the range \(-1 \leq \rho_{ij} \leq 1\) and is not unit dependent because the measurement units are scaled out in the construction of the ratio.

### 2.3 Percentiles and Value at Risk

The percentiles of a distribution are a set of summary statistics that summarise both the location and the spread of a distribution. Formally, a percentile is a measure that indicates the value below which a given percentage of observations in the sample fall. So the important measure of the location of a distribution, the median, below which 50\% of the observations of the random variable fall, is also the 50\(^{th}\) percentile. The median\(^4\) is an alternative to the sample mean as a measure of location and can be very important in financial distributions in which large outliers are encountered. The difference between the 25\(^{th}\) percentile (or first quartile) and the 75\(^{th}\) percentile (or third quartile) is known as the inter-quartile range, which provides an alternative to the standard deviation or variance as a measure of the dispersion of the distribution. It transpires that the percentiles of the distribution, particularly the 1\(^{st}\) and 5\(^{th}\) percentiles, are important statistics in the computation of a risk measure in finance known as value at risk.

Losses faced by financial institutions have the potential to be propagated through the financial system and undermine its stability. The onset of heightened fears for the riskiness of the banking system can be rapid and have widespread

---

\(^3\)This concept and terminology were also introduced by Sir Francis Galton, in this case independently of ideas developed by a French scientist, Auguste Bravais.

\(^4\)Yet another example of terminology introduced by Sir Francis Galton.
ramifications. The potential loss faced by banks is therefore a crucial measure of the stability of the financial sector. Pérignon and Smith (2010) examine the daily trading revenues, a measure of a bank’s fundamental soundness, for Bank of America. Summary measures and percentiles of the daily trading revenues from 2001 to 2004 are presented in Table 2.2.

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Percentiles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>1008</td>
</tr>
<tr>
<td>Mean</td>
<td>13.8699</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>14.9089</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.1205</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>4.9260</td>
</tr>
<tr>
<td>Maximum</td>
<td>84.3271</td>
</tr>
<tr>
<td>Minimum</td>
<td>−57.3886</td>
</tr>
</tbody>
</table>

A wave of banking collapses in the 1990s encouraged financial regulators to require banks to hold capital buffers against possible losses, following recommendations from the Basel Committee on Banking Supervision (1996) within the Bank of International Settlements (BIS). The mechanism for measuring bank exposure to possible losses employed a method called value at risk. Value at risk quantifies the loss that a bank can face on its trading portfolio within a given period and for a given confidence interval. In the context of a bank, value at risk is defined in terms of the lower tail of the distribution of trading revenues. Specifically, the 1% value at risk for the next $h$ periods conditional on information at time $T$ is the 1st percentile of expected trading revenue at the end of the next $h$ periods. For example, if the daily 1% $h$-period value at risk is $30$ million, then there is a 1% chance the bank will lose $30$ million or more. Although $30$ million is a loss, by convention the value at risk is quoted as a positive amount.

There are three common ways to compute value at risk.

(i) Historical Simulation

The historical method simply computes the percentiles of the empirical distribution from historical data. Based on the sample percentiles in Table 2.2 the 1% daily value at risk for Bank of America using all available historical data (2001 - 2004) is

$$\text{VaR} \left(1\%, \text{daily} \right) = 24.82 \text{ m}.$$  

(ii) The Variance Method
This method assumes that the trading revenues are normally distributed. Since 1% of the distribution lies in the tail delimited by $-2.33$ then, using the summary statistics reported in Table 2.2,

$$VaR(1\%, \text{daily}) = 13.8699 - 2.33 \times 14.9089 = 20.87\ m.$$  

This value is slightly lower than that provided by historical simulation because the assumption of normality ignores the slightly fatter tails exhibited by the empirical distribution of daily trading revenues.

(iii) Monte Carlo Simulation

The third method involves simulating a model for daily trading revenues several times and constructing simulated percentiles. This approach is revisited in Chapter 7.

![Figure 2.12: Time series plot of the daily trading revenue of Bank of America from 2 January 2001 to 31 December 2004. Also shown is the 1% historical value at risk of $24.82 million (dashed line) and the 1% value at risk actually reported by Bank of America on each day.](image)

Figure 2.12 plots the daily trading revenue of the Bank of America together with the 1% daily value at risk reported by the bank on each day. For comparative purposes the 1% historical value at risk is also shown. By construction, daily trading revenue will fall below the historical value at risk 1% of the time. During this period, however, the Bank of America had only four violations of the 1% daily reported value at risk. Since $4/1008 = 0.4\%$, it follows that during this period the Bank of America was over-conservative in its estimation of daily value at risk.
2.4 The Efficient Market Hypothesis

An important and controversial theory in finance is the efficient market hypothesis, which states in its most general form that all available information concerning the value of a risky asset is factored into the current price of the asset (Fama, 1965; Samuelson, 1965). There are many ways to test this theory and the literature on the subject is vast. Two simple ways to use sample statistics to examine this proposition are first to test asset returns for autocorrelation and second to compare the variance of asset returns over different time horizons.

2.4.1 Return Predictability

An implication of the efficient market hypothesis is that there is no predictable pattern in financial returns, that is they behave randomly over time. One way to test the predictability of returns is based on the autocorrelation statistic

\[ acf(k) = \frac{T^{-1} \sum_{t=k+1}^{T} (r_t - \bar{r})(r_{t-k} - \bar{r})}{T^{-1} \sum_{t=1}^{T} (r_t - \bar{r})^2}. \]  

(2.8)

This statistic measures the strength of association between the current return, \( r_t \), and the return on the same asset \( k \) periods earlier \( r_{t-k} \). In equation (2.8), the numerator represents the autocovariance of returns \( k \) periods apart and the denominator represents the variance of returns. If returns exhibit no autocorrelation then future movements in returns are unpredictable in terms of their own past history. But if returns exhibit positive or negative autocorrelation, then successive values of returns tend to have the same sign and this pattern can be exploited in predicting the future behaviour of returns.

Table 2.3 gives the first 10 autocorrelations of hourly Deutsche Mark (DM)/$ exchange rate returns in column 2. All the autocorrelations appear close to zero, suggesting that these exchange rate returns are not predictable using their own past history. The foreign exchange market is considered to be efficient in this sense, an interpretation that is limited by the fact that only the history of own past returns is considered in making the assessment of market efficiency.

Autocorrelations of returns reveal information about the temporal dependence properties of the levels of returns. Applying the same approach to squared returns gives

\[ acf^2(k) = \frac{T^{-1} \sum_{t=k+1}^{T} (r_t^2 - \bar{r}^2)(r_{t-k}^2 - \bar{r}^2)}{T^{-1} \sum_{t=1}^{T} (r_t^2 - \bar{r}^2)^2}, \]  

(2.9)
Table 2.3

Autocorrelation properties of returns and functions of returns for the hourly DM/$ exchange rate for the period 1 January 1986 00:00 to 15 July 1986 11:00.

| Lag | $r_t$ | $r_t^2$ | $|r_t|$ | $|r_t|^{0.5}$ |
|-----|-------|---------|--------|-----------|
| 1   | -0.0610 | 0.1340  | 0.2036 | 0.2003    |
| 2   | -0.0013 | 0.0845  | 0.1241 | 0.1296    |
| 3   | 0.0202  | 0.0164  | 0.0678 | 0.0710    |
| 4   | -0.0318 | 0.0512  | 0.0778 | 0.0650    |
| 5   | 0.0318  | 0.0327  | 0.0316 | 0.0224    |
| 6   | -0.0584 | 0.0093  | 0.0237 | 0.0310    |
| 7   | 0.0152  | -0.0258 | -0.0333| -0.0238   |
| 8   | -0.0182 | -0.0257 | -0.0127| 0.0064    |
| 9   | -0.0013 | -0.0297 | -0.0385| -0.0375   |
| 10  | 0.0190  | -0.0205 | -0.0377| -0.0338   |

where

$$r^2 = \frac{1}{T} \sum_{t=1}^{T} r_t^2.$$ 

This statistic reveals information about the autocorrelation properties of the squared levels of returns. Column 3 of Table 2.3 suggests that while the level of returns are not predictable, the same cannot be said of their squared level or variation. Note that this conclusion does not violate the efficient markets hypothesis, which is solely concerned with the expected value of the level of returns. The application of autocorrelations to squared returns represents an important diagnostic tool in models of time-varying volatility which is dealt with in Part III.

Autocorrelations can also be computed for various power transformations of returns, such as

$$r_t^3, \quad r_t^4, \quad |r_t|, \quad |r_t|^a.$$ 

The first two transformations provide evidence of autocorrelations in skewness and kurtosis respectively. The third transformation, which employs the modulus (or absolute magnitude) of returns, leads to an alternative measure of the presence of autocorrelation in the variance. The last case simply represents a general power transformation. For example, setting $a = 0.5$ computes the autocorrelation of the standard deviation (the square root of the variance). The presence of stronger autocorrelation in squared returns than returns themselves, suggests that other transformations of returns may reveal even stronger autocorrelation patterns and this conjecture is borne out by the results reported in Table 2.3.
2.4.2 The Variance Ratio

An alternative way to examine the efficient market hypothesis is to compare the variance on returns over different time horizons. Consider the variances of the 1-period returns and the n-period returns

\[
s_1^2 = \frac{1}{T} \sum_{t=1}^{T} (r_t - \bar{r})^2, \quad s_n^2 = \frac{1}{T} \sum_{t=1}^{T} (r_{nt} - n\bar{r})^2,
\]

in which

\[
r_t = \log P_t - \log P_{t-1} \\
r_{nt} = \log P_t - \log P_{t-n} = r_t + r_{t-1} + \cdots + r_{t-(n-1)},
\]

and \(n\bar{r}\) represents the sample mean of the n-period returns \(r_{nt}\).

If there is no autocorrelation the variance of n-period returns should equal \(n\) times the variance of the 1-period returns. The ratio

\[
VR_n = \frac{s_n^2}{n s_1^2},
\]

is known as the variance ratio and has the following implications for the properties of excess returns:

\[
VR_n = \begin{cases} 
1 & \text{[No autocorrelation]} \\
> 1 & \text{[Positive autocorrelation]} \\
< 1 & \text{[Negative autocorrelation]}
\end{cases}
\]

The first of these results is easily demonstrated. Consider an \(n = 3\) period return

\[
r_{3t} = r_t + r_{t-1} + r_{t-2},
\]

which is the sum of the three 1-period returns. Let the sample mean for the 1-period returns be \(\bar{r}\). Subtracting \(\bar{r}\) from both sides three times gives

\[
(r_{3t} - 3\bar{r}) = (r_t - \bar{r}) + (r_{t-1} - \bar{r}) + (r_{t-2} - \bar{r}).
\]
Squaring both sides and averaging over a sample of size \( T \) gives

\[
\begin{align*}
\sigma_3^2 &= \frac{1}{T} \sum_{t=1}^{T} (r_{3t} - 3\bar{r})^2 \\
&= \frac{1}{T} \sum_{t=1}^{T} (r_t - \bar{r})^2 \quad \text{[Variance of } r_t]\ \\
&+ \frac{1}{T} \sum_{t=1}^{T} (r_{t-1} - \bar{r})^2 \quad \text{[Variance of } r_{t-1}] \\
&+ \frac{1}{T} \sum_{t=1}^{T} (r_{t-2} - \bar{r})^2 \quad \text{[Variance of } r_{t-2}] \\
&+ \frac{2}{T} \sum_{t=1}^{T} (r_t - \bar{r})(r_{t-1} - \bar{r}) \quad \text{[Autocovariance of } r_t, r_{t-1}] \\
&+ \frac{2}{T} \sum_{t=1}^{T} (r_t - \bar{r})(r_{t-2} - \bar{r}) \quad \text{[Autocovariance of } r_t, r_{t-2}] \\
&+ \frac{2}{T} \sum_{t=1}^{T} (r_{t-1} - \bar{r})(r_{t-2} - \bar{r}) \quad \text{[Autocovariance of } r_{t-1}, r_{t-2}] \].
\end{align*}
\]

This expansion requires values for \( r_0 \) and \( r_{-1} \). To implement this formulation in practice, the summation ranges are suitably adjusted to the available data. In the case of zero sample autocovariances (or no sample autocorrelation) the relationship simplifies to

\[
\sigma_3^2 = \frac{1}{T} \sum_{t=1}^{T} (r_t - \bar{r})^2 + \frac{1}{T} \sum_{t=1}^{T} (r_{t-1} - \bar{r})^2 + \frac{1}{T} \sum_{t=1}^{T} (r_{t-2} - \bar{r})^2.
\]

Assuming that the sample variance for \( r_t \) is the same as the sample variance for \( r_{t-1} \) and \( r_{t-2} \) then gives

\[
\sigma_3^2 = 3\sigma_1^2,
\]

when there is no sample autocorrelation in the \( n = 3 \) period returns A more detailed discussion of the autocorrelation function is provided in Chapter 4. The assumption of equal variances for \( r_{t-1} \) and \( r_{t-2} \) falls under a general assumption known as (covariance) stationarity and is addressed in detail in Chapters 4 and 5. Modelling with variables that do not satisfy this assumption is dealt with in Chapter 6.

## 2.5 Exercises

The data required for the exercises are available for download as EViews workfiles (*.wf1), Stata datafiles (*.dta), comma delimited text files (*.csv) and as Excel spreadsheets (*.xlsx).

1. **Equity Prices, Dividends and Returns**
The data are monthly observations on United States equity prices and dividends for the period January 1871 to September 2016. For the subsample January 1950 to September 2016 do the following.

(a) Plot the equity price over time and interpret its time series properties.
(b) Plot the natural logarithm of the equity price over time and interpret its time series properties.
(c) Plot the return on equities over time and interpret its time series properties.
(d) Plot the price and dividend series using a line chart and comment on the results.
(e) Compute the dividend yield and plot this series using a line chart and comment on the results.
(f) Compare the graphs in parts (a) and (b) and discuss the time series properties of equity prices, dividend payments and dividend yields.
(g) The present value model predicts a linear relationship between the logarithm of equity prices and the logarithm of dividends. Use a scatter diagram to verify this property and comment on your results.
(h) For the returns on United States equities calculate the sample mean, variance, skewness and kurtosis. Interpret the statistics.

2. Yields

The data are monthly observations from December 1946 to February 1987 on United States zero coupon bond yields for maturities ranging from 2 months to 9 months.

(a) Plot the 2, 3, 4, 5, 6 and 9 months United States zero coupon yields using a line chart and comment on the results.
(b) Compute the spreads on the 3-month, 5-month and 9-month zero coupon yields relative to the 2-month yield. Plot these spreads using a line chart and comment on their properties.
(c) Compare the graphs in parts (a) and (b) and discuss the time series properties of yields and spreads.
3. Duration Times Between American Airline (AMR) Trades

- amr.wf1, amr.dta, amr.xlsx

The file contains high frequency data (one second intervals) giving a snapshot of the trades recorded by American Airlines (AMR) from 9.30am to 4.00pm on 1 August 2006.

(a) Use a histogram to graph the empirical distribution of the duration times between American Airline trades. Compare the graph with Figure 2.9.

(b) Interpret the shape of the distribution of duration times.

4. Exchange Rates

- hour.*

The data are hourly observations for the period 0.00am 1 January 1986 to 11.00am 15 July 1986 on the $/£ and $/DM exchange rates.

(a) Draw a line chart of the $/£ exchange rate and discuss its time series characteristics.

(b) Compute the log returns on the $/£ exchange rate. Draw a line chart of this series and discuss its time series characteristics.

(c) Compare the graphs in parts (a) and (b) and discuss the time series properties of exchange rates and exchange rate log returns.

(d) Use a histogram to graph the empirical distribution of the log returns on the $/£ exchange rate. Compare the graph with Figure 2.11.

(e) Compute the first 10 autocorrelations of the log returns, squared log returns, absolute log returns and the square root of the absolute value of log returns. Comment on the results.

(f) Repeat parts (a) to (e) using the $/DM exchange rate and comment on the time series characteristics, empirical distributions and patterns of autocorrelation for the two series. Discuss the implications of these results for the efficient market hypothesis.

5. Value-at-Risk
CHAPTER 2. PROPERTIES OF FINANCIAL DATA

bankamerica.*

The data are daily observations for the period 2 January 2001 to 31 December 2004 on the trading revenue and reported VaR values for Bank of America.

(a) Compute summary statistics and percentiles for the daily trading revenues of Bank of America. Compare the results with Table 2.2.

(b) Draw a histogram of the daily trading revenue and superimpose a normal distribution on top of the plot. What do you deduce about the distribution of the daily trading revenues?

(c) Plot the trading revenue together with the historical 1% VaR and the reported 1% VaR. Compare the results with Figure 2.12.

(d) Now assume that a weekly VaR is required. Repeat parts (a) to (c) for weekly trading revenues.

6. Solnik and Roulet Measure of Financial Integration

integration.*

The data consist of daily stock prices in United States dollars on 10 Asian stocks (China, Hong Kong, Indonesia, Japan, Malaysia, Philippines, Singapore, South Korea, Taiwan, Thailand) as well as the United States stock price, from 1 January 1997 to 27 May 2016.

(a) Compute the percentage log returns of all 11 stocks

\[ r_{it} = 100 \times \left( \log P_{it} - \log P_{i,t-1} \right), \]

where \( P_{it} \) is the stock price index. Plot the log returns and discuss the change in volatility over the sample period.

(b) Solnik and Roulet (2000) measure the change in financial integration over time using the standard deviation of cross-market returns based on

\[ \sigma_t^2 = \frac{1}{N-1} \sum_{i=1}^{N} (r_{it} - \tau_t)^2. \]

where a decrease in \( \sigma_t \) represents an improvement in financial market integration. Compute \( \sigma_t \) for the \( N = 10 \) Asian stocks and discuss the change in regional financial integration in Asia over the sample period.

(c) Use the Hodrick-Prescott filter to smooth the integration measure computed in part (b) for the \( N = 10 \) Asian stocks and discuss the change in financial integration in the region.
(d) Repeat parts (c) and (d) by extending the number of stocks to \( N = 11 \) by including United States log returns.

(e) An alternative measure of financial integration to \( \sigma_t \) is to compute the range at each point in time

\[
RANGE_t = \max_i r_{it} - \min_i r_{it}
\]

Using all Asian and United States stock returns \( N = 11 \) compute \( RANGE_t \) at each point in time and compare this measure of integration and the measure based on \( \sigma_t \).
Chapter 3
Linear Regression Models

One of the most important tools in empirical finance is linear regression. Its use is ubiquitous in financial econometrics and much more widely throughout the social and business sciences. The linear regression model provides a framework by which the movements of one financial variable are explained and predicted in terms of other variables called explanatory variables or covariates. One of the most important applications of linear regression is to the capital asset pricing model (CAPM), which provides a powerful way in which to model the risk of an asset from its exposure to the market. Multi-factor extensions of CAPM are also conveniently captured by extending the set of explanatory variables to contain size and growth factors as well as momentum.

The linear model can be developed and applied in a myriad of ways, including ranking the performance of portfolios and the construction of minimum variance portfolios. One important application is known as event analysis which provides a framework to capture qualitative effects of financial events by using explanatory variables known as dummy variables. Even in the case of simple linear regression, financial data produce fascinating complications that take us beyond standard textbook theory.

The treatment of linear regression presented here focusses more on its financial applications, with many of the theoretical results stated without detailed derivations. These derivations are widely available in standard introductory econometrics textbooks with some of the key results summarised for convenience in Appendices A and C.

---

1The concept of a regression line was invented by Sir Francis Galton (mentioned earlier in Chapter 2 in the context of the median, standard deviation, and correlation coefficient) in the course of his studies of sibling and parental heights. Among his many other accomplishments, Galton coined the terms regression and correlation and introduced the use of the letter r to represent the correlation coefficient, which was studied in Chapter 2. Versions of Galton’s correlation diagram for bivariate distributions still appear in introductory statistics texts.
3.1 The Capital Asset Pricing Model

The capital asset pricing model (CAPM) encapsulates the risk characteristics of an asset in terms of its so-called $\beta$-risk, which is given by the ratio

$$
\beta = \frac{\text{cov}(r_{it} - r_{ft}, r_{mt} - r_{ft})}{\text{var}(r_{mt} - r_{ft})}.
$$

(3.1)

This quantity is a measure of the exposure of the returns $r_{it}$ on the asset to movements in the market $r_{mt}$, relative to a risk-free rate of interest $r_{ft}$. Individual stocks, or even the portfolios of stocks, are classified as follows in terms of their degree of $\beta$-risk:

- Aggressive: $\beta > 1$
- Benchmark: $\beta = 1$
- Conservative: $0 < \beta < 1$
- Uncorrelated: $\beta = 0$
- Imperfect Hedge: $-1 < \beta < 0$
- Perfect Hedge: $\beta = -1$.

The returns on the market index such as the S&P500 for the United States is the usual standard that stocks are benchmarked against to measure their performance. Aggressive stocks on average move in excess to movements in the market. This is a particularly common characteristic of technology stocks. On the upside these stocks earn higher returns than the market in bull markets, but on the downside their average returns are expected to be lower. Benchmark stocks mimic the market so their movements tend to track the market one-for-one. Conservative stocks such as blue chip stocks also move in the same direction as the market, although their movements are attenuated. Unlike aggressive stocks the returns on conservative stocks are lower than the market on average during periods of market growth, whereas during periods of market declines the falls in returns of conservative stocks tend to be smaller than the falls experienced by the market.

Stocks that bear no market risk as they move independently of the market represent zero-beta stocks such as Treasury bonds and so by definition earn the risk-free rate of return. Hedge stocks such as gold tend to move in the opposite direction to the market thereby counter-balancing market movements. For an imperfectly hedged stock the absolute movements in the stock are less than the market movements, whereas for perfectly hedged stocks these movements match the movements in the market one-for-one on average, but in the opposite direction. Thus, a portfolio containing the market index and a perfectly hedged stock would bear no risk. Assuming an absence of arbitrage this portfolio would earn the risk-free rate.

The CAPM is formulated in terms of a linear regression model by expressing the relationship between the excess return on the asset $r_{it} - r_{ft}$ and the market
where the excess return on the asset represents the dependent variable and the excess return on the market represents the explanatory variable. The disturbance term \( u_t \) captures additional movements in the dependent variable not predicted by CAPM. Assuming that these additional movements are also independent of \( r_{mt} - r_{ft} \), then \( E[(r_{mt} - r_{ft})u_t] = 0 \). The regression model in (3.2) contains two unknown parameters. The first is the intercept parameter \( \alpha \) which captures the abnormal return to the asset over and above the asset’s exposure to the excess return on the market. The second is the slope parameter \( \beta \) which corresponds to the asset’s \( \beta \)-risk as defined in (3.1).

An important advantage of expressing the CAPM as a linear regression model is that it provides a convenient method of decomposing the total risk of an asset into systematic risk (or risk that is inherent to the entire market) from exposure to movements in the market, and idiosyncratic risk caused by other factors. Formally this decomposition is achieved by squaring both sides of (3.2) and then taking expectations

\[
E[(r_{it} - r_{ft})^2] = E[(\alpha + \beta (r_{mt} - r_{ft}))^2] + E(u_t^2),
\]

which uses the zero correlation property \( E[(r_{mt} - r_{ft})u_t] = 0 \). The systematic risk is also known as nondiversifiable risk while the idiosyncratic risk represents the diversifiable risk.

It is convenient to re-express the linear regression form of the CAPM in (3.2) by defining the dependent variable as \( y_t = r_{it} - r_{ft} \) and the explanatory variable as \( x_t = r_{mt} - r_{ft} \), so the CAPM is rewritten as

\[
y_t = \alpha + \beta x_t + u_t,
\]

in which \( \alpha \) and \( \beta \) are the \( \alpha \)-risk and \( \beta \)-risk of the asset \( x_t \), respectively. The disturbance term, \( u_t \), represents movements in \( y_t \) that are not explained by the model and it has the properties

\[
E(u_t) = 0, \quad E(u_t^2) = \sigma_u^2, \quad E(u_t u_{t-j}) = 0, \quad j \neq 0, \quad E(x_t u_t) = 0.
\]
property known as conditional heteroskedasticity. Techniques for relaxing the assumption of a constant $\sigma^2_u$ are discussed in Part III when the class of autoregressive conditional heteroskedasticity models (ARCH) is introduced. A test for time-varying variance is introduced later in the chapter.

The third property is that the movements in $u_t$ are uncorrelated with previous (or future) movements. If this is not the case the idiosyncratic component $u_t$ exhibits autocorrelation, a property that raises potential profit making (called arbitrage) opportunities because this information may be utilised to improve the predictions of future movements in $y_t$. The fourth and final property ensures that movements in the market ($x_t$) are not connected in any way with other factors (manifesting via $u_t$) caused by non-market movements. This assumption was stated earlier and also used in the decomposition of the total risk of an asset into its systematic and idiosyncratic risk components in (3.3).

To estimate the population risk parameters $\alpha$ and $\beta$ in the CAPM in (3.4) using data on $y_t$ and $x_t$ for $t = 1, 2, \cdots, T$, the linear regression approach chooses estimates that minimise the sum of squares function

$$S = \sum_{t=1}^{T} (y_t - \alpha - \beta x_t)^2. \quad (3.6)$$

The solution of this minimisation problem is demonstrated graphically in Figure 3.1 for the $\beta$-risk parameter. The estimate of $\beta$-risk, $\hat{\beta}$, is chosen here to minimise the function $S$ in (3.6) for some given $\alpha$. 

3.1. THE CAPITAL ASSET PRICING MODEL

Figure 3.1: Illustrating the ordinary least squares estimator of the parameter $\beta$ in the simple regression model for the case of given $\alpha$. The estimate $\hat{\beta}$ is the argument which minimises the function $S(\beta)$ given this value of $\alpha$.

Formally the full solution for both parameters $\alpha$ and $\beta$ is obtained by differentiating the sum of squares function $S$ with respect to $\alpha$ and $\beta$. Setting the derivatives to zero and rearranging these expressions yields the ordinary least squares estimators

$$
\beta\text{-risk: } \hat{\beta} = \frac{T^{-1} \sum_{t=1}^{T} (y_t - \bar{y})(x_t - \bar{x})}{T^{-1} \sum_{t=1}^{T} (x_t - \bar{x})^2} \tag{3.7}
$$

$$
\alpha\text{-risk: } \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}
$$

where $\bar{y} = T^{-1} \sum_{t=1}^{T} y_t$ and $\bar{x} = T^{-1} \sum_{t=1}^{T} x_t$ are the sample means. The standard errors of these estimators are, respectively,

$$
\text{se}(\hat{\alpha}) = \sqrt{\hat{\sigma}^2 \frac{T \sum_{t=1}^{T} x_t^2}{T \sum_{t=1}^{T} (x_t - \bar{x})^2}}, \quad \text{se}(\hat{\beta}) = \sqrt{\hat{\sigma}^2 \frac{T \sum_{t=1}^{T} x_t^2}{T \sum_{t=1}^{T} (x_t - \bar{x})^2}}, \tag{3.8}
$$

in which

$$
\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta} x_t)^2,
$$

is the estimate of the residual variance corresponding to a measure of the idiosyncratic risk of the asset. In computing the residual variance the degrees of freedom correction $T - 2$ is often used in the denominator of the expression. The standard errors are a measure of the precision of the estimators –
the smaller is the standard error the more precise is the least squares estimator of the population parameter. The properties of ordinary least squares estimators will be discussed more formally in Section 3.3.

The sample counterpart of the CAPM regression model in (3.4) is obtained by replacing the population parameters by their respective estimators as

$$y_t = \hat{\alpha} + \hat{\beta}x_t + \hat{u}_t. \quad (3.9)$$

The term $\hat{\alpha} + \hat{\beta}x_t$ represents the regression line, while $\hat{u}_t$ is known as the regression residual which provides an estimator of the idiosyncratic risk of the asset at time $t$. An important property of the least squares estimators that immediately follows from solving (3.6) is that the residual sum of squares

$$RSS = \sum_{t=1}^{T} \hat{u}_t^2, \quad (3.10)$$

is a minimum. It is for this reason that the least squares regression line is commonly referred to as the line of best fit.

The connection between the usual definition of $\beta$-risk given in (3.1) and the least squares estimator given in (3.7) is made transparent by noting that the numerator of $\hat{\beta}$ in (3.7) is the sample covariance between the excess returns on the asset and the market, while the denominator is the sample variance of the market. This is exactly the same estimator that would follow by directly replacing the population quantities in (3.1) with their sample counterparts, a result which establishes the link between $\beta$-risk and the linear regression model in this context. Finally, the estimator of $\alpha$-risk, $\hat{\alpha}$, is obtained by computing the difference between the average return on the asset over the sample period $\bar{y}$ and the average return implied by the CAPM model, namely $\beta\bar{x}$. If the estimate $\hat{\alpha} > 0$ the asset earns higher (abnormal) returns in excess of the return predicted by CAPM, while the opposite occurs if abnormal returns are lower and $\hat{\alpha} < 0$.

To illustrate the estimation of the CAPM, a data set comprising monthly returns to 10 United States industry portfolios for the period January 1927 to December 2013 is used, together with a benchmark of monthly returns to the market and a risk free rate of interest. The industry portfolios are: Consumer Nondurables, Consumer Durables, Manufacturing, Energy, Technology, Telecommunications, Wholesale and Retail, Healthcare, Utilities and Other, which is a catch-all portfolio that includes mining, construction, entertainment and finance. The return on the market is constructed as the value-weighted return of all CRSP firms incorporated in the United States and listed on the NYSE, AMEX, or NASDAQ and the risk free rate is the 1-month United States Treasury Bill rate.

The computation of the ordinary least squares estimates of the $\alpha$-risk and $\beta$-risk parameters is illustrated in Table 3.1 using the Nondurables portfolio.
The first 2 columns give respectively the monthly excess returns for the Non-durables portfolio, \( y_t \), and the market, \( x_t \). The least squares estimates are

\[
\hat{\beta} = \frac{T^{-1} \sum_{t=1}^{T} (x_t - \overline{x})(y_t - \overline{y})}{T^{-1} \sum_{t=1}^{T} (x_t - \overline{x})^2} = \frac{23273.5059/1044}{30712.9719/1044} = 0.7578,
\]

\[
\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} = 0.6942 - 0.757 \times 0.6449 = 0.2055,
\]

and the estimate of the residual variance is

\[
\hat{\sigma}^2_u = \frac{1}{T - 2} \sum_{t=1}^{T} (y_t - \hat{\alpha} - \hat{\beta}x_t)^2 = \frac{1}{T - 2} \sum_{t=1}^{T} \hat{u}_t^2 = \frac{5026.2404}{1044 - 2} = 4.8236,
\]

where the degrees of freedom adjustment is used in the computation. The standard errors of \( \hat{\alpha} \) and \( \hat{\beta} \) are, respectively,

\[
\text{se}(\hat{\alpha}) = \sqrt{\frac{\hat{\sigma}^2_u \sum_{t=1}^{T} x_t^2}{T \sum_{t=1}^{T} (x_t - \overline{x})^2}} = \sqrt{\frac{4.8236 \times 31147.2117}{1044 \times 30712.9719}} = \sqrt{0.0047} = 0.0684,
\]

\[
\text{se}(\hat{\beta}) = \sqrt{\frac{\hat{\sigma}^2_u \sum_{t=1}^{T} x_t^2}{\sum_{t=1}^{T} (x_t - \overline{x})^2}} = \sqrt{\frac{4.8236}{30712.9719}} = \sqrt{0.0002} = 0.0125.
\]

The computations are repeated in Appendix C using the matrix formulation of the linear regression model.

The results obtained by estimating the CAPM for all of the 10 industry portfolios are given in Table 3.2. The aggressive portfolios (\( \hat{\beta} > 1 \)) are Durables, Manufacturing, Technology and Other. The remaining six portfolios are conservative portfolios (\( 0 < \hat{\beta} < 1 \)). As expected none of the industry portfolios provide a perfect hedge against systematic risk. All the portfolios indicate positive abnormal profits (\( \hat{\alpha} > 0 \)) with the exception of the Other portfolio where the estimate of \( \alpha \)-risk is \(-0.103\).

### 3.2 A Multi-factor CAPM

The CAPM has been extended in a number of ways to allow for additional determinants of excess returns. In a seminal paper, Fama and French (1993) augment the CAPM by including two additional risk factors to explain the return on a risky investment. These factors are: the performance of small stocks relative to big stocks (SMB), known as a Size factor; and the performance of value stocks relative to growth stocks (HML), known as a Value factor. In addition, Carhart (1997) suggests a further extension based on momentum (MOM) which captures the returns to a portfolio constructed by buying stocks with high returns over the past three to twelve months and selling...
Table 3.1: Calculations to estimate the CAPM for the Nondurables portfolio by ordinary least squares. The data are monthly from January 1927 to December 2013, where $y_t$ is the excess return on the Nondurables portfolio, $x_t$ is the excess return on the market and $\hat{u}_t$ is the least squares residual.

<table>
<thead>
<tr>
<th>Month</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$y_t$</td>
</tr>
<tr>
<td>0.3152 &amp; 4.6873 &amp; 4.2345</td>
<td></td>
</tr>
<tr>
<td>1.2779 &amp; 6.9129 &amp; 2.4727</td>
<td></td>
</tr>
<tr>
<td>1.3846 &amp; 1.1428 &amp; 1.4270 &amp; 1.4270</td>
<td></td>
</tr>
<tr>
<td>1.1147 &amp; 6.8083 &amp; 1.1151 &amp; 1.1151</td>
<td></td>
</tr>
<tr>
<td>2.1671 &amp; 1.1218 &amp; 1.5658 &amp; 0.9342</td>
<td></td>
</tr>
<tr>
<td>1.5690 &amp; 2.0505 &amp; 0.5095 &amp; 1.1151</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>2013 Jan.</td>
<td>2.0466</td>
</tr>
<tr>
<td>2.0600 &amp; 0.6896 &amp; 2.7242 &amp; 2.6849</td>
<td></td>
</tr>
<tr>
<td>3.8583 &amp; 1.0160 &amp; 0.0600 &amp; 0.3031</td>
<td></td>
</tr>
<tr>
<td>2.5205 &amp; 0.0160 &amp; 0.0160 &amp; 0.0160</td>
<td></td>
</tr>
<tr>
<td>2.4682</td>
<td>0.9160 &amp; 0.0160 &amp; 0.0160</td>
</tr>
<tr>
<td>1.1658</td>
<td>0.9160 &amp; 0.0160 &amp; 0.0160</td>
</tr>
<tr>
<td>2.2462</td>
<td>1.1658 &amp; 0.0160 &amp; 0.0160</td>
</tr>
<tr>
<td>1.927 Jan.</td>
<td>0.0521</td>
</tr>
<tr>
<td>1.0435 &amp; 6.6950 &amp; 1.1685 &amp; 0.0160</td>
<td></td>
</tr>
</tbody>
</table>

Note: Calculations to estimate the CAPM for the Nondurables portfolio by ordinary least squares. The data are monthly from January 1927 to December 2013, where $y_t$ is the excess return on the Nondurables portfolio, $x_t$ is the excess return on the market and $\hat{u}_t$ is the least squares residual.
Table 3.2

Ordinary least squares estimates of the CAPM in equation (3.2) using 10 industry portfolios using monthly data for the United States beginning January 1927 and ending December 2013. Standard errors are given in parentheses. The measures $R^2$ and $\hat{R}^2$ are the coefficient of determination and adjusted coefficient of determination, which are given in (3.17) and (3.18) and discussed later in this Chapter.

<table>
<thead>
<tr>
<th>Industry</th>
<th>$\hat{\alpha}$</th>
<th>$\hat{\beta}$</th>
<th>$R^2$</th>
<th>$\hat{R}^2$</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nondurables</td>
<td>0.205</td>
<td>0.758</td>
<td>0.778</td>
<td>0.778</td>
<td>5026.2</td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durables</td>
<td>0.003</td>
<td>1.244</td>
<td>0.747</td>
<td>0.747</td>
<td>16110.2</td>
</tr>
<tr>
<td></td>
<td>(0.123)</td>
<td>(0.022)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manufacturing</td>
<td>0.008</td>
<td>1.128</td>
<td>0.924</td>
<td>0.923</td>
<td>3234.9</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy</td>
<td>0.231</td>
<td>0.856</td>
<td>0.595</td>
<td>0.595</td>
<td>15289.8</td>
</tr>
<tr>
<td></td>
<td>(0.119)</td>
<td>(0.022)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tech</td>
<td>0.009</td>
<td>1.236</td>
<td>0.825</td>
<td>0.825</td>
<td>9939.8</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.022)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telecom</td>
<td>0.152</td>
<td>0.657</td>
<td>0.591</td>
<td>0.591</td>
<td>9176.5</td>
</tr>
<tr>
<td></td>
<td>(0.092)</td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail</td>
<td>0.107</td>
<td>0.969</td>
<td>0.789</td>
<td>0.788</td>
<td>7734.7</td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.016)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health</td>
<td>0.255</td>
<td>0.841</td>
<td>0.650</td>
<td>0.650</td>
<td>11696.9</td>
</tr>
<tr>
<td></td>
<td>(0.104)</td>
<td>(0.019)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Utilities</td>
<td>0.089</td>
<td>0.782</td>
<td>0.576</td>
<td>0.576</td>
<td>13805.6</td>
</tr>
<tr>
<td></td>
<td>(0.113)</td>
<td>(0.021)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>−0.103</td>
<td>1.126</td>
<td>0.876</td>
<td>0.876</td>
<td>5524.7</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

stocks with low returns over the same period. This factor captures the presence of herd behaviour among investors who are following market movements. The market factor of CAPM and these three additional factors are plotted in Figure 3.2 for the United States from January 1927 to December 2013.

The CAPM in (3.2) is extended to allow for the additional size, value and momentum factors by expressing the multi-factor CAPM as a multiple regression model (that is, a regression model with multiple regressors). Letting $EMKT_t$ be the excess return on the market factor which is $(r_{mt} - r_{ft})$ in (3.2), $SMB_t$ and $HML_t$ represent the Fama-French size and growth factors and $MOM_t$ be the momentum factor, the multi-factor CAPM is specified as

$$r_{it} - r_{ft} = \alpha + \beta_1 EMKT_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 MOM_t + u_t,$$

(3.11)

where $u_t$, as before, is a disturbance term. The term $\alpha + \beta_1 EMKT_t$ on the right
Figure 3.2: Monthly data for market, size, value and momentum factors of the extended CAPM model for the period January 1927 to December 2012.

The multiple regression representation of the multi-factor CAPM in (3.11) is generalised by redefining $y_t$ as the dependent variable and $x_{1t}, x_{2t}, \ldots, x_{kt}$ as a set of $K$ explanatory variables representing all of the factors. The multiple linear regression model is

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_K x_{kt} + u_t,$$

which contains $K + 1$ unknown parameters. The disturbance term $u_t$ is as-
3.3. **PROPERTIES OF ORDINARY LEAST SQUARES**

Estimation of the $K + 1$ unknown parameters in (3.12) proceeds as before by redefining the sum of squares function in (3.6) as

$$S = \sum_{t=1}^{T} (y_t - \alpha - \beta_1 x_{1t} - \beta_2 x_{2t} - \cdots - \beta_K x_{Kt})^2. \quad (3.13)$$

Minimising this expression with respect to the unknown $K + 1$ parameters yields a solution which is the multivariate analogue of the bivariate solution given in (3.7). In dealing with the multiple regression model it is convenient to use matrix notation. An introduction to the general linear regression model in matrix notation is given in Appendix C.

The results obtained by estimating the multi-factor CAPM for the 10 industry portfolios are given in Table 3.3. A comparison of the $\alpha$-risk estimates ($\hat{\alpha}$) and the beta risk estimates ($\hat{\beta}_1$) obtained from the multi-factor CAPM model and the estimates based on the CAPM presented in Table 3.2 show that the two sets of parameter estimates are qualitatively very similar in terms of signs and magnitudes. This result also suggests that the role of the market factor is still very important in pricing assets as its influence is undiminished through the addition of other factors into the model. Inspection of the signs on the parameter estimates of the size ($\hat{\beta}_2$), value ($\hat{\beta}_3$) and momentum ($\hat{\beta}_4$) factors reveals that different industries have vastly differing exposures to these factors.

### 3.3 Properties of Ordinary Least Squares

The ordinary least squares estimators of the parameters of the CAPM in (3.4) and more generally the multiple linear regression model in (3.12), have a number of useful properties in both small and large samples. As large samples are of particular interest in financial econometrics given the availability of high-frequency large datasets, only these properties of the ordinary least squares estimators are discussed here. In particular, the properties considered are called asymptotic properties and these correspond to the (theoretical) situation where sample sizes progressively increase to become infinitely large so that $T \to \infty$. Asymptotic analysis of this type is particularly valuable because it enables us to understand the properties of fitting regression equations in the ideal situation of having an infinite amount of data. In such situations we naturally expect good performance in regression estimation.

Three important asymptotic properties are presented below which describe such performance of the estimators in relation to the respective population parameters. The first property focuses on the probability of an estimator being arbitrarily close to the true parameter value (consistency), which describes the capacity to locate the true parameter when there is an infinite amount of data. The second concerns the variance of its distribution (asymptotic effi-
Ordinary least squares estimates of the multi-factor CAPM using 10 industry portfolios using monthly data for the United States beginning January 1927 and ending December 2013. Standard errors are given in parentheses.

<table>
<thead>
<tr>
<th>Industry</th>
<th>( \alpha )</th>
<th>( \beta_1 )</th>
<th>( \beta_2 )</th>
<th>( \beta_3 )</th>
<th>( \beta_4 )</th>
<th>( R^2 )</th>
<th>( R^2 )</th>
<th>RSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nondurables</td>
<td>0.179</td>
<td>0.767</td>
<td>-0.032</td>
<td>0.028</td>
<td>0.025</td>
<td>0.779</td>
<td>0.778</td>
<td>5001.6</td>
</tr>
<tr>
<td>Duraives</td>
<td>0.089</td>
<td>1.172</td>
<td>0.034</td>
<td>0.145</td>
<td>-0.152</td>
<td>0.763</td>
<td>0.762</td>
<td>15098.7</td>
</tr>
<tr>
<td>Manuf</td>
<td>-0.012</td>
<td>1.104</td>
<td>-0.001</td>
<td>0.128</td>
<td>-0.021</td>
<td>0.929</td>
<td>0.929</td>
<td>2994.7</td>
</tr>
<tr>
<td>Energy</td>
<td>0.072</td>
<td>0.895</td>
<td>-0.210</td>
<td>0.265</td>
<td>0.116</td>
<td>0.627</td>
<td>0.626</td>
<td>14083.4</td>
</tr>
<tr>
<td>Tech</td>
<td>0.190</td>
<td>1.242</td>
<td>0.098</td>
<td>-0.372</td>
<td>-0.089</td>
<td>0.853</td>
<td>0.852</td>
<td>8380.0</td>
</tr>
<tr>
<td>Telecom</td>
<td>0.258</td>
<td>0.678</td>
<td>-0.140</td>
<td>-0.095</td>
<td>-0.070</td>
<td>0.606</td>
<td>0.604</td>
<td>8846.6</td>
</tr>
<tr>
<td>Retail</td>
<td>0.170</td>
<td>0.960</td>
<td>0.075</td>
<td>-0.121</td>
<td>-0.039</td>
<td>0.794</td>
<td>0.794</td>
<td>7527.0</td>
</tr>
<tr>
<td>Health</td>
<td>0.297</td>
<td>0.893</td>
<td>-0.105</td>
<td>-0.166</td>
<td>0.023</td>
<td>0.666</td>
<td>0.665</td>
<td>11172.6</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.007</td>
<td>0.777</td>
<td>-0.180</td>
<td>0.311</td>
<td>0.008</td>
<td>0.620</td>
<td>0.618</td>
<td>12386.6</td>
</tr>
<tr>
<td>Other</td>
<td>-0.141</td>
<td>1.043</td>
<td>0.064</td>
<td>0.335</td>
<td>-0.081</td>
<td>0.918</td>
<td>0.918</td>
<td>3635.7</td>
</tr>
</tbody>
</table>

Consistency

The least squares estimators of \( \alpha \)-risk, \( \hat{\alpha} \), and \( \beta \)-risk, \( \hat{\beta} \), are based on extracting information from the data over the period \( t = 1, 2, \ldots, T \). If the assumptions on the disturbance term of the multiple linear regression model in (3.5) remain satisfied as more and more information is employed by increasing \( T \), these estimators approach the population alpha and \( \beta \)-risk parameters, \( \alpha \) and \( \beta \), in the limit as \( T \to \infty \). This property is known as consistency and shows that closer estimates of the population parameters are typically obtained by having more information.

A more formal statement of consistency of the two risk estimators is

\[
\text{plim}(\hat{\alpha}) = \alpha, \quad \text{plim}(\hat{\beta}) = \beta, \tag{3.14}
\]

where the use of the term plim denotes a limit that is taken in probability. The use of a probability limit rather than a deterministic limit (denoted as lim) is
in recognition that $\hat{\beta}$ is a random quantity, being a function of data which is itself random. Even though $\hat{\beta}$ approaches $\beta$ as $T$ increases, the inherent randomness of $\hat{\beta}$ ensures that the path to $\beta$ remains random as $T$ increases, unlike that of a convergent deterministic sequence. In recognition of this characteristic, convergence is said to occur in probability as $T \to \infty$.

Figure 3.3 illustrates the consistency of the estimates of $\hat{\alpha}$ and $\hat{\beta}$ in the CAPM regression. The CAPM is simulated with population parameters of $\alpha = 1$ and $\beta = 1$, and sample sizes ranging from $T = 10$ to $T = 10000$. The excess returns on the market, $x_t$, are simulated from the standard normal distribution $x_t \sim N(0, 1)$, while the disturbance term $u_t$ is simulated from the normal distribution $u_t \sim N(0, 0.25)$. For each sample the estimates of $\alpha$ and $\beta$ are plotted in the figure. The property of consistency is clearly demonstrated with both $\hat{\alpha}$ and $\hat{\beta}$ approaching their true values of 1, as the sample size progressively increases. This figure also demonstrates the random character of the estimators and the random nature of the convergence towards the population values. Even for very large samples of $T = 10000$, random fluctuations in the path occur, but these continue to contract as $T$ increases, revealing the nature of the probabilistic convergence.

Figure 3.3: Illustrating the convergence of the least squares estimates of $\alpha$ and $\beta$ in the CAPM regression (3.4). Estimates of $\alpha$ (left panel) and $\beta$ (right panel) are shown for sample sizes from 10 to 10000 in increments of 1 observation.
Asymptotic Efficiency

In discussing the property of consistency with the aid of Figure 3.3, it was observed that the parameter estimates from the simulation experiment did not monotonically converge towards the population parameters as $T \to \infty$. Instead, each estimate continued to exhibit random fluctuations about the true population parameter but with its randomness diminishing as $T$ increased. A measure of the spread of this randomness in estimation is captured by estimating the asymptotic variance.

The asymptotic variances of the ordinary least squares estimators with respect to their population parameters in (3.4) are given by the following expressions

$$\text{var}(\hat{\alpha}) = E[ (\hat{\alpha} - \alpha)^2 ] = \frac{\sigma_u^2 \sigma_x^2 + \mu_x^2}{\sigma_x^2} \frac{1}{T},$$

$$\text{var}(\hat{\beta}) = E[ (\hat{\beta} - \beta)^2 ] = \frac{\sigma_u^2 \sigma_x^1}{\sigma_x^2} \frac{1}{T},$$

(3.15)

where $\sigma_u^2$ is the variance of the disturbance term $u_t$, and $\mu_x = E(x)$ and $\sigma_x^2 = E[(x - \mu_x)^2]$ are, respectively, the mean and variance of the explanatory variable $x$. An important feature of these expressions is that the asymptotic variances are inversely related to the sample size $T$. Since $\sigma_u^2$, $\sigma_x^2$ and $\mu_x^2$, are all constants by definition, increasing the sample size progressively diminishes the spread of the estimators around their population parameters. In fact, doubling the sample size decreases the asymptotic variances by 50%. This property of the asymptotic variances is demonstrated by the simulation experiment presented in Figure 3.3. In the simulation experiments $\mu_x = 0$, $\sigma_x^2 = 1$ and $\sigma_u^2 = 0.25$, so the asymptotic variances in (3.15) are calculated to be

$$\text{var}(\hat{\alpha}) = \frac{\sigma_u^2 \sigma_x^2 + \mu_x^2}{\sigma_x^2} \frac{1}{T} = \frac{0.25 \cdot 1 + 0^2}{1} \frac{1}{T} = \frac{0.25}{T},$$

$$\text{var}(\hat{\beta}) = \frac{\sigma_u^2 \sigma_x^1}{\sigma_x^2} \frac{1}{T} = \frac{0.25 \cdot 1}{1} \frac{1}{T} = \frac{0.25}{T},$$

with the spread of the parameter estimates progressively falling as the sample size progressively increases.

An important property of the ordinary least squares estimator is that under the assumptions described above and certain mild technical conditions, the asymptotic variances given in (3.15) are the smallest possible variances attainable by any consistent estimator. The ordinary least squares estimator is therefore said to be asymptotically efficient.
Asymptotic Normality

The first and second properties of the least squares estimators $\hat{\alpha}$ and $\hat{\beta}$ are concerned with the mean and the variance of the asymptotic distributions of these estimators. The third and final property is concerned with the shape of the asymptotic distribution. Under the same conditions described above the asymptotic distributions of $\hat{\alpha}$ and $\hat{\beta}$, are normal with means given by (3.14) and variances given by (3.15). Upon appropriate scaling of the estimation errors by $\sqrt{T}$, the asymptotic normality property of the least squares estimators is formalised by writing

$$\sqrt{T}(\hat{\alpha} - \alpha) \overset{d}{\to} N \left( 0, \frac{\sigma_\epsilon^2 (\sigma_x^2 + \mu_x^2)}{\sigma_x^2} \right), \quad \sqrt{T}(\hat{\beta} - \beta) \overset{d}{\to} N \left( 0, \frac{\sigma_x^2 \sigma_\epsilon^2}{\sigma_x^2} \right),$$

(3.16)

where the symbol $\overset{d}{\to}$ represents convergence in distribution as $T \to \infty$, signifying that the sequence of distributions converges to a well defined limit distribution, which in this case is a normal distribution.

Figure 3.4: Illustrating asymptotic normality of the least squares estimator, $\hat{\beta}$, in the regression model (3.4). Estimates of the densities of the distribution of $\sqrt{T}(\hat{\beta} - 1)/0.5$ for sample sizes 10 and 50 are computed using 10000 replications in each case. Also shown is the standard normal distribution.

Figure 3.4 illustrates the asymptotic normality property of the least squares $\beta$-risk estimator $\hat{\beta}$. Using the same simulation setup as before for a fixed $x_t$ series, the asymptotic distributions are presented by using (3.16) to construct
the standardised statistic

\[
\sqrt{T} \frac{\hat{\beta} - \beta}{\sqrt{\sigma_u^2 / \sigma_x^2}} = \sqrt{T} \frac{\hat{\beta} - 1}{0.5},
\]

which has a standard normal asymptotic distribution, \(N(0,1)\). In Figure 3.4 two sampling distributions are presented for samples of size \(T = 10\) and \(T = 50\), and shown against the asymptotic \(N(0,1)\) distribution. These distributions are all centred on zero as predicted by the consistency property (3.14). Even the sampling distribution based on samples of size \(T = 10\) is reasonably close to the asymptotic \(N(0,1)\) distribution. Increasing the sample size to \(T = 50\) delivers an extremely close correspondence with the asymptotic distribution.

3.4 Diagnostics

The estimated CAPM and its linear regression model representation in (3.9) are based on the assumption that this model is correctly specified and corresponds to the process that generated the data used to fit the regression. But if asset returns are determined by several factors as in the multi-factor CAPM (3.11) then the 1-factor CAPM misspecifies the true generating mechanism of the data. To test this fundamental assumption a number of diagnostic procedures are available. These may be classified into three categories involving diagnostics on the dependent variable \(y_t\), the explanatory variables \(\{x_{1t}, x_{2t}, \cdots, x_{Kt}\}\), and the disturbance term \(u_t\).

3.4.1 Diagnostics on the Dependent Variable

A natural measure of the success of an estimated model is given by the proportion of the variation in the dependent variable explained by the model. This measure is called the coefficient of determination and is defined by the ratio

\[
R^2 = \frac{\text{Explained sum of squares}}{\text{Total sum of squares}} = \frac{\sum_{t=1}^{T} (y_t - \bar{y})^2 - \sum_{t=1}^{T} \hat{u}_t^2}{\sum_{t=1}^{T} (y_t - \bar{y})^2}.
\]

The coefficient of determination satisfies the inequality \(0 \leq R^2 \leq 1\), with values close to unity suggesting a good model fit and values close to zero representing a poor fit. Given the decomposition of the total risk of an asset into systematic and idiosyncratic components in (3.3), this expression suggests that \(R^2\) provides an estimate of the proportion of the total risk that is
due to systematic risk, with the proportion attributable to idiosyncratic risk estimated by $1 - R^2$.

The $R^2$ statistics for the estimated CAPM applied to the 10 portfolios are given in Table 3.2. The Manufacturing portfolio is estimated to have the highest proportion of systematic risk with a $R^2$ of 0.924, with the Other portfolio having the second highest with an $R^2$ of 0.876. The Utilities portfolio has the lowest estimated proportion of systematic risk with the split between systematic and idiosyncratic being roughly 50-50.

A related measure to the $R^2$ statistic is the adjusted coefficient of determination

$$R^2 = 1 - (1 - R^2) \frac{T - 1}{T - K - 1}, \quad (3.18)$$

which penalises the $R^2$ statistic for specifying a model with additional parameters arising from the inclusion of more explanatory variables. For the multi-factor CAPM in (3.11) the additional factors are the size, growth and momentum factors. If these additional factors contribute further (that is, beyond the effect of the penalty) to explaining movements in the excess returns on an asset then $R^2$ increases, whereas if they do not then $R^2$ falls. However, in both scenarios $R^2$ increases which implies that the $R^2$ of the multi-factor CAPM is always greater than the $R^2$ computed for the CAPM. This result is confirmed by comparing the $R^2$ statistics in Table 3.2 for the CAPM which are all lower than the corresponding statistics in Table 3.3 for the multi-factor CAPM. A comparison of the $R^2$ across the two tables identifies the relative importance of the additional factors included in the multi-factor CAPM. Inspection of these statistics shows evidence in favour of the 4-factor CAPM with $R^2$ increasing for all portfolios with the exception of the Nondurables portfolio where it does not change.

### 3.4.2 Diagnostics on the Explanatory Variables

The CAPM in (3.2) explains the price of a risky asset in terms of its exposure to systematic risk. To establish the significance of these risk factors, statistical tests are now performed on the estimated parameters of the CAPM.

#### Single Parameter Tests

To test the importance of the market factor in the CAPM regression equation, the $\beta$-risk parameter estimate is tested to assess whether the true parameter is zero using a $t$ test. The null and alternative hypotheses are

$$H_0 : \beta = 0 \quad \text{[market factor not priced]}$$
$$H_1 : \beta \neq 0 \quad \text{[market factor priced]}.$$
The $t$ statistic to perform this test is given by the ratio
\[ t = \frac{\hat{\beta}}{se(\hat{\beta})}, \] (3.19)
where $\hat{\beta}$ is the estimated coefficient of $\beta$ and $se(\hat{\beta})$ is the corresponding standard error. Note that in computing the standard errors, the estimate of the residual variance $\hat{\sigma}_u^2$ is computed using the degrees of freedom correction $T - K - 1$ to take account of the $K + 1$ unknown parameters in the regression model. The test statistic in (3.19) has an asymptotic normal distribution under the null hypothesis. In small samples and under certain conditions (including normally distributed errors $u_t$) this statistic has a Student $t$ distribution with $T - K - 1$ degrees of freedom.$^2$

For a 5% significance level the decision rule is
\begin{align*}
\text{p value} < 0.05 : & \text{ Reject } H_0 \text{ at the 5\% level of significance} \\
\text{p value} > 0.05 : & \text{ Fail to reject } H_0 \text{ at the 5\% level of significance},
\end{align*} (3.20)
where the $p$ value is the probability calculated under the null hypothesis of finding a more extreme value of the test statistic than the computed one. From Table 3.2, the $t$ statistic on the coefficient representing $\beta$-risk in the Nondurables portfolio is
\[ t = \frac{\hat{\beta}}{se(\hat{\beta})} = \frac{0.7578}{0.0125} = 60.47. \]

This $t$ statistic yields a $p$ value of $0.000 < 0.05$ showing that the excess return on the market is significant at the 5% level in determining movements in the excess return on the Nondurables portfolio. Inspection of the $t$ statistics for the remaining portfolios reveals qualitatively similar results with the excess market return being significant in all cases. This result provides strong support for the applicability of the CAPM in all sectors of the United States economy.

The application of a $t$ test to determine the strength of the $\beta$-risk of an asset can also be applied to test for the asset’s $\alpha$-risk by testing the hypothesis $\alpha = 0$. More generally, $t$ tests can be constructed for other types of hypotheses such as testing whether an asset benchmarks the market by testing the hypothesis $\beta = 1$. In this case the $t$ statistic in (3.19) becomes
\[ t = \frac{\hat{\beta} - 1}{se(\hat{\beta})}. \] (3.21)

Applications of these types of $t$ tests are developed in the exercises.

$^2$The Student $t$ distribution was developed by the English statistician William Gosset in 1908 who wrote under the pen name Student. He worked in the Guinness brewery and on farms in Ireland, where only small samples of data were available on different varieties of barley which he was testing for effectiveness in production. In his honour, the process of taking the ratio (3.19) of an estimate to its standard error is now universally known as studentisation.
Joint Parameter Tests

In many financial applications, it is of interest to perform tests of hypotheses on several parameters jointly. In the case of the multi-factor CAPM in equation (3.11), a joint test of the significance of all of the additional factors is based on the null and alternative hypotheses

\[ H_0 : \beta_2 = \beta_3 = \beta_4 = 0 \quad \text{[CAPM preferred]} \]

\[ H_1 : \text{at least one restriction fails} \quad \text{[Multi-factor CAPM preferred]}, \]

which corresponds to \( R = 3 \) restrictions to be tested. Under the null hypothesis the restrictions imply that CAPM is the preferred model whereas under the alternative hypothesis the more general multi-factor CAPM is preferred.

The general form of the joint test statistic involves estimating the model twice. First, under the null hypothesis in which the restrictions are imposed; and second under the alternative hypothesis in which the restrictions are relaxed. Letting \( RSS_0 \) represent the residual sum of squares from the restricted model as given by CAPM, and \( RSS_1 \) the residual sum of squares from the unrestricted model as given by the multi-factor CAPM, the test statistic takes the form

\[
J = \frac{(RSS_0 - RSS_1)}{RSS_1 / (T - K - 1)},
\]

where \( T \) is the sample size and \( K \) is the number of factors in the unrestricted model which for the multi-factor CAPM is \( K = 4 \). If the restrictions are consistent with the true model there will be little difference in the two sums of squares resulting in a small value of the test statistic in (3.22). However, if the restrictions are not valid then imposing the restrictions typically leads to a large increase in \( RSS_0 \) relative to \( RSS_1 \) resulting in a large value of the test statistic. Distinguishing between the null and alternative hypothesis is achieved formally by assessing the significance of the statistic. Significance is established by using the property that the test statistic in (3.22) has a chi-square distribution with \( R \) degrees of freedom (written \( \chi^2_R \)) in large samples, a property that holds under conditions covered by the assumptions already made. This distribution is used to generate critical values and \( p \) values of the test statistic.

In small samples the statistic \( F = J / R \) has an F distribution with \( R \) and \( T - K - 1 \) degrees of freedom (written \( F_{R,T−K−1} \)). Like the \( t \) distribution, the \( F \) distribution is the exact small sample distribution of the ratio under the assumption that the data are normally distributed.\(^5\)

\(^3\)A critical value is a cut-off point in the distribution that determines whether the observed value of the statistic is significant – or extreme enough – to reject the null hypothesis.

\(^4\)The \( p \) value of a test is the calculated probability of the observed or a more extreme value of the statistic when the null hypothesis is true.

\(^5\)This distribution is named after the English statistician Sir Ronald Fisher (1890-1962), one of the founders of modern experimental agricultural research, who developed many new statistical methods to analyze experimental data at the Rothamsted Experimental Station located in the country town of Harpenden near London.
To implement the test of the multi-factor CAPM for the Nondurables portfolio the restricted residual sum of squares is $RSS_0 = 5026.2$ from the column of Table 3.2 and the unrestricted sum of squares is $RSS_1 = 5001.6$ from the last column of Table 3.3. Substituting these values into (3.22) together with $T = 1044$ and $K = 4$ yields a statistic of

$$J = \frac{(RSS_0 - RSS_1)}{RSS_1/(T - K - 1)} = \frac{(5026.2 - 5001.6)}{5001.6/(1044 - 4 - 1)} = 5.11.$$ 

Under the null hypothesis $J$ is distributed asymptotically as $\chi^2_3$ with a $p$ value of 0.1640. This result provides statistical support for the one-factor CAPM suggesting that the additional risk factors given by the growth, value, and momentum factors are not priced for this portfolio. The opposite result occurs however, for all of the remaining industry portfolios with the joint test providing strong support for the multi-factor CAPM at the 5% level. This result is provided as an exercise.

### 3.4.3 Diagnostics on the Disturbance Term

The third set of diagnostic tests concerns the properties of the disturbance term $u_t$. For the CAPM regression model to be specified correctly there should be no systematic and useful information about excess returns remaining in the disturbance term. Otherwise there will be arbitrage opportunities which can be used to improve predictions of the dependent variable (excess returns) and thereby enhance profits. This property suggests that model specification tests can be based on the following set of hypotheses

$$H_0 : u_t \text{ is random} \quad [\text{model is specified correctly}]$$

$$H_1 : u_t \text{ is non-random} \quad [\text{model is misspecified}].$$

The adoption of tests concerning $u_t$ is especially important for those situations where the coefficient of determination is found to be extremely small. Even though a low $R^2$ suggests that the estimated model provides a poor fit of the dependent variable, this does not necessarily mean that the model is misspecified as the data could be extremely noisy as is often the case with high frequency data on financial returns. An $R^2$ lower than 5% is not at all uncommon in such regressions as financial returns are notoriously difficult to predict. Indeed, from an efficient markets point of view, there may be no effective predictor of future returns. In this scenario the estimated model can nonetheless still be viewed as well specified provided that the null hypothesis that $u_t$ is random is not rejected.

#### Testing for Autocorrelation

The aim of a test for residual autocorrelation is to detect the presence of temporal dependence in the disturbance terms. The null and alternative hypothe-
ses are respectively
\[ H_0 : \text{No autocorrelation in the disturbances} \]
\[ H_1 : \text{Autocorrelation in the disturbances.} \]

If there is no autocorrelation this provides support for the model, whereas rejection of the null hypothesis suggests that the model excludes important information that may aid prediction. In the presence of autocorrelation the ordinary least squares estimator is no longer efficient.

The autocorrelation test is based on the correlation between the least squares residuals \( \hat{u}_t \) and past residuals obtained by specifying the following multiple regression equation

\[
\hat{u}_t = \gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{2t} + \cdots + \gamma_K x_{K,t} + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \cdots + \rho_p \hat{u}_{t-p} + v_t,
\]

where \( v_t \) is a disturbance term. The autocorrelation in the residuals is captured by the regressors \( \{ \hat{u}_{t-1}, \hat{u}_{t-2}, \cdots, \hat{u}_{t-p} \} \) which allow for autocorrelation for up to \( p \) lags in this specification. The choice of \( p \) may be informed by theory or guided by the frequency of the data. For example, in monthly data \( p \) might be set to 12. A test of the null hypothesis of no autocorrelation is performed by testing the joint restrictions \( \rho_1 = \rho_2 = \cdots = \rho_p = 0 \). The test statistic of autocorrelation of order \( p \), is given by

\[
AR(p) = TR^2,
\]

where \( T \) is the sample size and \( R^2 \) is the coefficient of determination from estimating (3.23) by ordinary least squares. Under the null hypothesis \( AR(p) \) has a \( \chi^2_p \) distribution, which is used in determining significance and the outcome of the test.

A test for first order autocorrelation of the residuals of the estimated CAPM for the Nondurables portfolio yields a statistic of 22.646. Using the \( \chi^2_1 \) distribution the \( p \) value is 0.000 showing strong rejection of the null hypothesis in favour of autocorrelation. The presence of autocorrelation in the residuals may be the result of the exclusion of factors that are also important in pricing the portfolio such as the additional factors included in the multi-factor CAPM. Alternatively, the problem may arise because important dynamics such as lagged dependent and explanatory variables have been incorrectly excluded. Extensions of the linear regression model to allow for dynamic relationships are discussed in Chapter 4.

**Testing for Heteroskedasticity**

An important assumption of the linear regression model given in (3.5) is that the variance of the disturbance term \( \sigma^2_u \) is assumed to be constant over the sample. In the context of the CAPM this assumption implies that the (squared)
idiosyncratic risk does not vary over time. As discussed earlier, constant variance is known as homoskedasticity whereas non-constant variance is known as heteroskedasticity. To test this assumption the null and alternative hypotheses are respectively

\[ H_0 : \text{Homoskedasticity} \quad [\sigma^2_u \text{ is constant}] \]
\[ H_1 : \text{Heteroskedasticity} \quad [\sigma^2_u \text{ is time-varying}] \]

Just as in the case of autocorrelation, the presence of heteroskedasticity in the residuals will mean that the ordinary least squares estimator is no longer efficient.

A test for heteroskedasticity is based on determining whether the squared residuals \( \hat{u}_t^2 \) are influenced by the explanatory variables in the regression model. For the CAPM the test is constructed by first specifying the following regression equation

\[
\hat{u}_t^2 = \gamma_0 + \gamma_1 x_1 t + \gamma_2 x_2^2 t + v_t, \tag{3.25}
\]

where \( v_t \) is a disturbance term, and then testing the joint restrictions \( \gamma_1 = \gamma_2 = 0 \). The test statistic for heteroskedasticity is

\[
\text{HETERO} = TR^2, \tag{3.26}
\]

where \( T \) is the sample size and \( R^2 \) is the coefficient of determination from estimating (3.25) by ordinary least squares. This test is also known as White’s test of heteroskedasticity (White, 1980). Under the null hypothesis \( \text{HETERO} \) has a \( \chi^2_2 \) distribution. Although the discussion here focuses on the case of only a single regressor, the test is easily generalised to accommodate multiple regressors.

**Testing for ARCH**

Figure 3.5 provides a plot of the squared residuals from estimating the CAPM for the Nondurables portfolio. Performing the White test yields a test statistic of 34.990. The associated \( p \) value is 0.000 providing strong evidence against the null hypothesis of a constant disturbance variance. One possible reason for this rejection is the presence of volatility clustering whereby large (small) movements are associated with further large (small) movements. This phenomenon is formally known as autoregressive conditional heteroskedasticity (ARCH), a concept that was originally proposed by Engle (1982).

To perform a test for ARCH in the disturbance variance, the null and alternative hypotheses are

\[ H_0 : \text{[No ARCH]} \]
\[ H_1 : \text{[ARCH]} \]

A test for ARCH of arbitrary order \( p \) is performed by estimating the following regression equation

\[
\hat{u}_t^2 = \gamma_0 + \gamma_1 \hat{u}_{t-1}^2 + \gamma_2 \hat{u}_{t-2}^2 + \cdots + \gamma_p \hat{u}_{t-p}^2 + v_t, \tag{3.27}
\]
where \( v_t \) is a disturbance term. To establish the presence of ARCH it is usually only necessary to run the test for small values of \( p \). The test statistic is

\[
ARCH(p) = TR^2,
\]

where \( T \) is the sample size and \( R^2 \) comes from the estimating equation (3.27) by ordinary least squares. Under the null hypothesis ARCH(\( p \)) has a \( \chi^2_p \) distribution.

A test of first order ARCH by setting \( p = 1 \) in (3.27) for the Nondurables portfolio generates a test statistic of 51.911 with a \( p \) value of 0.000. This strong rejection of the No ARCH null hypothesis is not surprising given the strong visual evidence of volatility clustering already identified in Figure 3.5. This empirical result also suggests that the CAPM linear regression specification needs to be extended to allow for a time-varying disturbance variance.

**Testing for Normality**

The diagnostic tests discussed so far are concerned with either the mean of the linear regression model (autocorrelation) or its variance (White and ARCH). The Jarque-Bera statistic (Jarque and Bera, 1987) provides a test of normality of the disturbance distribution. Even though the assumptions given in (3.5) to establish the properties of the least squares estimator do not require the
assumption of normality, the properties of the diagnostic tests presented are nonetheless affected by the underlying disturbance distribution.

The null and alternative hypotheses of the test of normality are, respectively,

\[ H_0 : u_t \text{ is normally distributed} \]
\[ H_1 : u_t \text{ is nonnormally distributed.} \]

The test statistic is

\[ JB = T \left( \frac{SK^2}{6} + \frac{(KT - 3)^2}{24} \right), \quad (3.29) \]

where \( T \) is the sample size, and \( SK \) and \( KT \) are skewness and kurtosis measures, respectively, of the least squares residuals

\[ SK = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\hat{u}_t}{\hat{u}_u} \right)^3, \quad KT = \frac{1}{T} \sum_{t=1}^{T} \left( \frac{\hat{u}_t}{\hat{u}_u} \right)^4. \]

Effectively the JB statistic is a joint test for the presence of skewness and excess kurtosis relative to the normal distribution. Under the null hypothesis the JB statistic is asymptotically distributed as \( \chi^2_2 \).

The value of the JB statistic of the residuals from the CAPM using the Non-
durables portfolio is \( JB = 192.8 \). Using the \( \chi^2_2 \) distribution, the \( p \) value is 0.000, resulting in strong rejection of the null hypothesis at the 5% level. An implication of nonnormality is that it affects the standard errors of the parameter estimates resulting in incorrect inferences.

In cases where the null hypothesis is rejected in favour of nonnormal disturbances three potential solutions are available. The first solution is to try to address the problem at its source. One possible option is to use indicator variables to capture extreme observations which are creating the nonnormality problem (see Exercise 7), thereby effectively removing those observations from being covered in the regression model. The second is to model the nonnormality directly by replacing the normal distribution with a more appropriate alternative, a solution that is explored in Chapter 10 when using maximum likelihood estimation. The third approach is to do nothing in terms of estimation but to adjust the estimates of the standard errors that are used in statistical testing to take account of the presence of nonnormality in the data, a solution that is based on quasi maximum likelihood estimation and discussed in Section 10.6.

### 3.5 Measuring Portfolio Performance

The performance of a portfolio is commonly measured in terms of its expected return in excess of the return from a risk-free asset relative to the risk
of the portfolio. Three well-known measures of a portfolio’s performance are

\[
S = \frac{\mu_p - r_f}{\sigma_p},
\]

Treynor index: \( T = \frac{\mu_p - r_f}{\beta} \)

Jensen’s alpha: \( \alpha = \mu_p - r_f - \beta(\mu_m - r_f) \)

where \( \mu_p \) and \( \mu_m \) are respectively the expected returns on the portfolio and the market, \( r_f \) is the risk-free rate, and the risk measures are portfolio risk \( \sigma_p \) and \( \beta \)-risk \( \beta \) from CAPM. The Sharpe ratio \( S \) demonstrates how well the return of an asset compensates the investor for the risk taken, which is measured as the average excess return to the portfolio per unit of total portfolio risk (Sharpe, 1966). The Treynor index \( T \) is similar to the Sharpe ratio except that the risk of the portfolio is defined in terms of systematic risk using the \( \beta \)-risk from the CAPM (Treynor, 1966). Jensen’s alpha \( \alpha \) is the abnormal return on a portfolio relative to its systematic risk which is obtained from the CAPM regression (Jensen, 1968). Of the three measures Jensen’s alpha is possibly the most widely used as a positive \( \alpha \) is a necessary condition for good performance.

Table 3.4
Performance measures and rankings for 10 United States industry portfolios computed using monthly returns data for the period January 1927 to December 2013.

<table>
<thead>
<tr>
<th>Industry</th>
<th>Mean ( \hat{\mu}_p )</th>
<th>Std. Dev. ( \hat{\sigma}_p )</th>
<th>Sharpe Ratio</th>
<th>Treynor Index</th>
<th>Jensen’s Alpha</th>
<th>Rank ( S )</th>
<th>Rank ( T )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nondur.</td>
<td>0.98</td>
<td>4.66</td>
<td>0.15</td>
<td>0.92</td>
<td>0.21</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Durables</td>
<td>1.09</td>
<td>7.79</td>
<td>0.10</td>
<td>0.65</td>
<td>0.00</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Manuf.</td>
<td>1.02</td>
<td>6.36</td>
<td>0.12</td>
<td>0.65</td>
<td>0.01</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Energy</td>
<td>1.07</td>
<td>6.01</td>
<td>0.13</td>
<td>0.92</td>
<td>0.23</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Tech.</td>
<td>1.09</td>
<td>7.37</td>
<td>0.11</td>
<td>0.65</td>
<td>0.01</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Telecom.</td>
<td>0.86</td>
<td>4.64</td>
<td>0.12</td>
<td>0.88</td>
<td>0.15</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Retail</td>
<td>1.02</td>
<td>5.91</td>
<td>0.12</td>
<td>0.76</td>
<td>0.11</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Health</td>
<td>1.09</td>
<td>5.66</td>
<td>0.14</td>
<td>0.95</td>
<td>0.26</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Utilities</td>
<td>0.88</td>
<td>5.59</td>
<td>0.11</td>
<td>0.76</td>
<td>0.09</td>
<td>8</td>
<td>5</td>
</tr>
<tr>
<td>Other</td>
<td>0.91</td>
<td>6.52</td>
<td>0.10</td>
<td>0.55</td>
<td>-0.10</td>
<td>10</td>
<td>10</td>
</tr>
</tbody>
</table>

The relative performance rankings of the 10 portfolios used in the CAPM application are presented in Table 3.4, which are computed by replacing the population parameters in (3.30) with their corresponding sample quantities. The estimates of the expected portfolio returns and risk are given in columns 2 and 3 of the table as \( \hat{\mu}_p \) and \( \hat{\sigma}_p \). To complete the calculations the expected excess return on the market is estimated as \( \hat{\mu}_m - r_f = 0.6449 \), where the risk-free rate is taken as the sample average \( r_f = 0.2873 \). In the case of the Non-
durables portfolio the three performance measures are computed as

\[
\begin{align*}
\text{Sharpe:} & \quad \hat{S} = \frac{\hat{\mu}_p - r_f}{\hat{\sigma}_p} = \frac{0.9814 - 0.2873}{4.6608} = 0.1489 \\
\text{Treynor:} & \quad \hat{T} = \frac{\hat{\mu}_p - r_f}{\hat{\beta}} = \frac{0.9814 - 0.2873}{0.7577} = 0.9161 \\
\text{Jensen:} & \quad \hat{\alpha} = \hat{\mu}_p - r_f - \hat{\beta}(\bar{\mu}_m - r_f) \\
& \quad = 0.9814 - 0.2873 - 0.7577 \times (0.6449) = 0.2054,
\end{align*}
\]

with \( \hat{\beta} \) taken from the CAPM regressions in Table 3.2. The Health portfolio has the highest rankings being ranked number one by the Treynor index and Jensen’s alpha, and second by the Sharpe ratio. The Nondurables portfolio is also highly ranked by all three measures with a top ranking using the Sharpe ratio, second using the Treynor index and third by Jensen’s alpha. The worst performing portfolio is Others which is ranked last by all three measures. Despite some broad consistency in the rankings of many of the portfolios, the ranking are clearly not perfectly consistent, a phenomenon that is commonly encountered in practical performance evaluation.

Indeed, different rankings among these measures is to be expected as the Sharpe ratio, for instance, accounts for total portfolio risk whereas the Treynor measure adjusts excess portfolio returns for systematic risk only. However, some similarity between the rankings provided by Treynor’s index and Jensen’s alpha is also unsurprising given that the alpha measure is derived from a CAPM regression which explicitly accounts for systematic risk via the inclusion of the market factor.

### 3.6 Minimum Variance Portfolios

A common goal in choosing a portfolio of assets is to minimise the overall risk of the portfolio, as measured by its variance, or squared volatility. The minimum variance portfolio problem can be formulated neatly in terms of linear regression where many of the key statistics have precise interpretations.

To illustrate the problem of constructing a minimum variance portfolio in a two asset case, monthly equity prices for the United States stocks Microsoft and Walmart are used.

To derive the minimum variance portfolio, consider a portfolio consisting of two assets with returns \( r_{1t} \) and \( r_{2t} \) which satisfy

\[
\begin{align*}
\text{Mean:} & \quad \mu_1 = E[r_{1t}], \quad \mu_2 = E[r_{2t}], \\
\text{Variance:} & \quad \sigma_1^2 = E[(r_{1t} - \mu_1)^2], \quad \sigma_2^2 = E[(r_{2t} - \mu_2)^2], \\
\text{Covariance:} & \quad \sigma_{12} = E[(r_{1t} - \mu_1)(r_{2t} - \mu_2)].
\end{align*}
\]
The return on the portfolio is
\[ r_{pt} = w_1 r_{1t} + w_2 r_{2t}, \quad w_1 + w_2 = 1, \] (3.31)
in which \( w_1 \) and \( w_2 \) are the weights that define the relative contributions of each asset in the portfolio.

The expected return on this portfolio is
\[ \mu_p = E(w_1 r_{1t} + w_2 r_{2t}) = w_1 E(r_{1t}) + w_2 E(r_{2t}) = w_1 \mu_1 + w_2 \mu_2, \] (3.32)
which is a weighted sum of the expected returns on the two assets. A measure of the portfolio risk is given by the variance
\[ \sigma^2_p = E[(r_{pt} - \mu_p)^2] = w_1^2 \sigma^2_1 + w_2^2 \sigma^2_2 + 2w_1w_2 \sigma_{12}. \] (3.33)

Using the adding-up restriction in (3.31), the risk of the portfolio is re-expressed in terms of the portfolio weight on the first asset as
\[ \sigma^2_p = w_1^2 \sigma^2_1 + (1 - w_1)^2 \sigma^2_2 + 2w_1(1 - w_1)\sigma_{12}. \] (3.34)

To find the optimal portfolio that minimises risk, the following optimisation problem is solved
\[ \min_{w_1} \sigma^2_p. \]
Differentiating (3.34) with respect to \( w_1 \), setting the derivative to zero and solving for \( w_1 \) gives the minimum variance weights
\[ w_1 = \frac{\sigma^2_2 - \sigma_{12}}{\sigma^2_1 + \sigma^2_2 - 2\sigma_{12}}, \quad w_2 = \frac{\sigma^2_1 - \sigma_{12}}{\sigma^2_1 + \sigma^2_2 - 2\sigma_{12}}, \] (3.35)
where the solution for \( w_2 \) uses the adding-up condition \( w_2 = 1 - w_1 \). The minimum variance weights are a function of the returns variances \( \sigma^2_1 \) and \( \sigma^2_2 \) and the covariance \( \sigma_{12} \).

Figure 3.6 contains the log returns on Microsoft and Walmart for the period April 1990 to July 2004. The sample averages and covariance matrix of the returns are, respectively,
\[ \hat{\mu}_1 = 0.020877, \quad \hat{\mu}_2 = 0.013496, \quad \begin{bmatrix} \hat{\sigma}^2_1 & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}^2_2 \end{bmatrix} = \begin{bmatrix} 0.011333 & 0.002380 \\ 0.002380 & 0.005759 \end{bmatrix}. \]

In computing the elements of the covariance matrix the biased form, presented in Chapter 2, is used in which \( T \) appears in the denominator instead of \( T - 1 \) in the sample variance matrix. To compute the minimum variance portfolio using the data on Microsoft and Walmart, the weights in (3.35) are estimated by replacing the population parameters by their sample quantities. The minimum variance weights are
\[ \hat{w}_1 = \frac{\hat{\sigma}^2_2 - \hat{\sigma}_{12}}{\hat{\sigma}^2_1 + \hat{\sigma}^2_2 - 2\hat{\sigma}_{12}}. \]
\[ \hat{w}_2 = 1 - \hat{w}_1 = 1 - 0.274 = 0.726. \]  

Thus, the optimal portfolio to minimise risk is to allocate 0.274 of the investor’s wealth to Microsoft and 0.726 to Walmart.

The average return on the minimum variance portfolio is estimated as

\[ \hat{\mu}_p = \hat{w}_1 \hat{\mu}_1 + \hat{w}_2 \hat{\mu}_2 = 0.274 \times 0.020877 + 0.726 \times 0.013496 = 0.015519. \]  

An estimate of the risk of the minimum variance portfolio is

\[ \hat{\sigma}^2_p = \hat{w}_2^2 \hat{\sigma}^2_1 + (1 - \hat{w}_1)^2 \hat{\sigma}^2_2 + 2 \hat{w}_1 (1 - \hat{w}_1) \hat{\sigma}_{12} \]
\[ = 0.274^2 \times 0.011333 + (1 - 0.274)^2 \times 0.005759 \]
\[ + 2 \times 0.274 \times (1 - 0.274) \times 0.002380 \]
\[ = 0.004833. \]  

Comparing this estimate to the individual risks on Microsoft and Walmart shows that the risk on the portfolio is indeed reduced.

To explore the relationship between the linear regression model in (3.2) and the minimum variance portfolio consider the following regression equation

\[ r_{2t} = \alpha + \beta (r_{2t} - r_{1t}) + u_t, \]

where \( y_t = r_{2t} \) is the dependent variable corresponding to the return on Walmart, and the explanatory variable is defined as \( x_t = r_{2t} - r_{1t} \), the excess return between Microsoft and Walmart. Using the log returns data on Microsoft and Walmart in Figure 3.6, the sample statistics of \( y_t \) and \( x_t \) are

\[ \bar{y} = 0.013496, \quad \bar{x} = -0.007380, \quad \hat{\sigma}^2_x = 0.012331, \quad \hat{\sigma}_{yx} = 0.003379. \]

The least squares parameter estimates are

\[ \hat{\beta} = \frac{\hat{\sigma}_{yx}}{\hat{\sigma}^2_x} = \frac{0.003379}{0.012331} = 0.274, \]
\[ \hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} = 0.013496 - 0.274 \times (-0.007380) = 0.015519. \]

A comparison of the slope parameter estimate of \( \hat{\beta} = 0.274 \) and the estimate of the minimum variance weight on Microsoft based on the finance calculations in (3.36) shows that they are equal. In addition, the estimate of the intercept \( \hat{\alpha} = 0.015519 \) also equals the estimate of the average return on the portfolio given in (3.37).

The standard error of the estimated regression equation is \( \hat{\sigma}_u = 0.069931 \) while the variance is \( \hat{\sigma}^2_u = 0.069931^2 = 4.8903 \times 10^{-3} \). This value is similar to the estimate of the risk of the minimum variance portfolio given in (3.38),
3.7 EVENT ANALYSIS

Figure 3.6: The returns to United States stocks Microsoft and Walmart computed using monthly data for the period April 1990 to July 2004.

but it is not the same. The difference between the two risk estimates is simply due to the degrees of freedom correction used to compute \( \hat{\sigma}_u^2 \). Readjusting this estimate as

\[
\frac{T - 2}{T} \hat{\sigma}_u^2 = \frac{171 - 2}{171} \times 0.069931^2 = 0.004833,
\]

now yields exactly the same value reported using the minimum variance calculations.

3.7 Event Analysis

Event analysis is widely used in empirical finance to model the effects of qualitative changes on financial variables arising from a discrete event. Typical events that are relevant in finance arise from announcements such as the change in a company’s CEO, an antitrust decision, a monetary policy announcement, or dramatic news events. In undertaking an event study the overall event is decomposed into three sub-events: the part that is anticipated
by the market, the part that occurs at the time of the event, and the part that happens after the event has occurred.

To model qualitative effects in an event analysis an indicator variable is defined which takes the value 1 if the event occurs at a point in time \( t \) and 0 for a non-occurrence of the event, namely

\[
I_t = \begin{cases} 
1 & \text{[event]} \\
0 & \text{[non-event]}
\end{cases}
\]  

(3.39)

This indicator variable is also commonly referred to as a dummy variable. A typical event study involves specifying a regression equation based on some given model that represents normal market returns, and then modifying this equation to account for designated events through the inclusion of indicator variables as defined in (3.39). The modifications involve the inclusion of separate dummy variables at each point in time over the event window to capture the abnormal returns, whether these may be positive or negative. The parameter on a particular dummy measures the abnormal return associated with that event, representing the return over and above the normal return.

In December of 2005, Lee Raymond retired as the CEO of Exxon receiving, at the time, the largest retirement package ever recorded of around US$400m. To determine how the market viewed this event, an event study is performed on the returns on Exxon around the time the announcement is made. A multiple regression model with a 5-month event window is specified as

\[
\begin{align*}
\begin{aligned}
   r_t & = \beta_0 + \beta_1 r_{mt} \\
        & + \delta_{-2} I_{Oct,t} + \delta_{-1} I_{Nov,t} + \delta_0 I_{Dec,t} + \delta_1 I_{Jan,t} + \delta_2 I_{Feb,t} + u_t,
\end{aligned}
\end{align*}
\]  

(3.40)

The indicator variables are defined around the event in December of 2005 as

- **Pre-event:**
  \[
  I_{Oct,t} = \begin{cases} 
1 & \text{Oct. 2005} \\
0 & \text{Otherwise}
\end{cases}, \quad I_{Nov,t} = \begin{cases} 
1 & \text{Nov. 2005} \\
0 & \text{Otherwise}
\end{cases}
\]

- **Actual event:**
  \[
  I_{Dec,t} = \begin{cases} 
1 & \text{Dec. 2005} \\
0 & \text{Otherwise}
\end{cases}
\]

- **Post-event:**
  \[
  I_{Jan,t} = \begin{cases} 
1 & \text{Jan. 2006} \\
0 & \text{Otherwise}
\end{cases}, \quad I_{Feb,t} = \begin{cases} 
1 & \text{Feb. 2006} \\
0 & \text{Otherwise}
\end{cases}
\]

The abnormal return in the month of the announcement is \( \delta_0 \); in the months prior to the announcement, the abnormal returns are given by \( \delta_{-2} \) and \( \delta_{-1} \); and for the months after the announcement they are given by \( \delta_1 \) and \( \delta_2 \).

Using returns on Exxon \((r_t)\) and the market \((r_{mt})\), the multiple regression model in (3.40) is estimated from January 1970 to February 2006. As the sample period stops in February 2006 the event window corresponds to the last 5
3.8. EXERCISES

sample observations. The estimated model with standard errors in parentheses is

\[
\begin{align*}
    r_t &= 0.009 + 0.651 r_{mt} - 0.121 I_{\text{Oct},t} + 0.007 I_{\text{Nov},t} \\
    & \quad - 0.041 I_{\text{Dec},t} + 0.086 I_{\text{Jan},t} - 0.059 I_{\text{Feb},t} + \bar{u}_t.
\end{align*}
\]

Inspection of the parameter estimates and standard errors suggests that the market not only anticipated the event in December 2005 two months earlier where the parameter estimate on the October indicator variable is statistically significant, but that it viewed the retirement package unfavourably with the return on Exxon falling by 0.121 in October. This result is also supported by the fact that on the day that the announcement is made in December of 2005, the estimated return of −0.041 is also negative, although not statistically significant. However, there appears to be a market correction in the month following the announcement with a statistically significant increase in the return on Exxon of 0.086 in January of 2006.

The net effect of the retirement package on the market is negative with the total abnormal return equalling

\[
\text{Total} = -0.121 + 0.007 - 0.041 + 0.086 - 0.059 = -0.128.
\]

The significance of this value is determined by performing a joint test of the hypothesis that the parameters on the 5 event indicator variables in (3.40) are zero. The null and alternative hypotheses are

\[
\begin{align*}
    H_0 &: \quad \delta_{-2} = \delta_{-1} = \delta_0 = \delta_1 = \delta_2 = 0 \quad \text{[Normal returns]} \\
    H_1 &: \quad \text{at least one restriction is not valid} \quad \text{[Abnormal returns]}
\end{align*}
\]

Under the null hypothesis that \( \delta_{-2} = \delta_{-1} = \delta_0 = \delta_1 = \delta_2 = 0 \), the regression model in (3.40) reduces to the market model. Using a chi square test the statistic is 16.399. As there are 5 restrictions being tested this statistic has a p value of 0.0058. These findings suggest that the net effect of the retirement package was to dampen the returns on Exxon and that the dampening effect was statistically significant.

3.8 Exercises

The data required for the exercises are available for download as EViews workfiles (*.wfl), Stata datafiles (*.dta), comma delimited text files (*.csv) and as Excel spreadsheets (*.xlsx).

1. The CAPM
CHAPTER 3. LINEAR REGRESSION MODELS

The data are monthly observations for the period April 1990 to July 2004 on the equity prices of Exxon, General Electric, IBM and Microsoft and Walmart, together with the price of Gold, the S&P 500 index and a short-term interest rate.

(a) Compute the monthly excess returns on Exxon, General Electric, Gold, IBM, Microsoft and Walmart. Be particularly careful when computing the correct risk free rate to use. [Hint: the variable TBILL is quoted as an annual rate.]

(b) Letting \( r_{it} - r_{ft} \) represent the excess return on asset \( i \) and \( r_{mt} - r_{ft} \) represent the excess return on the market, estimate the CAPM

\[
 r_{it} - r_{ft} = \alpha + \beta_1 (r_{mt} - r_{ft}) + u_t,
\]

for each asset where \( u_t \) is a disturbance term. Interpret the estimated \( \beta \) - and \( \alpha \)-risks.

(c) For each asset, test the restrictions \( \alpha = 0 \) and \( \beta = 1 \) individually and jointly. Interpret the results of the test.

2. Fama-French Three Factor Model

The data set contains the monthly Fama-French data for market, risk free, size, book-to-market and momentum factors for the period January 1927 to December 2013. The return on the market is constructed as the value-weighted return of all CRSP firms incorporated in the United States and listed on the NYSE, AMEX, or NASDAQ and the risk free rate is the 1-month United States Treasury Bill rate. The file also contains the monthly returns to 25 United States portfolios formed by sorting on size and book-to-market. The data are from Ken French’s webpage, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

(a) Estimate the Fama-French three factor model

\[
 r_{it} - r_{ft} = \alpha + \beta_1 EMKT_t + \beta_2 SMB_t + \beta_3 HML_t + u_t,
\]

for each of the 25 portfolios where \( EMKT_t \) is the excess return on the market, \( SMB_t \) is the size risk factor, \( HML_t \) is the value factor and \( u_t \) is a disturbance term. Interpret the parameter estimates on all three risk factors.
(b) Repeat part (a) for the 1-factor CAPM by imposing the restrictions $\beta_2 = \beta_3 = 0$. Interpret the estimated $\beta$-risk and compare with the estimates obtained in part (a).

(c) Perform a joint test of the size (SMB) and value (HML) risk factors in explaining excess returns in each portfolio.

3. **Present Value Model**

The data file contains monthly United States data on equity prices, $P_t$, and dividend payments, $D_t$, for the period January 1871 to September 2016.

(a) Estimate the present value model

$$p_t = \alpha + \beta d_t + u_t,$$

where $p_t = \log P_t$, $d_t = \log D_t$ and $u_t$ is a disturbance term.

(b) Examine the properties of the estimated model by performing the following diagnostic tests.

i. Plot the ordinary least squares residuals and interpret their time series patterns.

ii. Test for autocorrelation of orders 1 to 6.

iii. Test for ARCH.

iv. Test for normality.

(c) Estimate the implied discount factor used to compute the present value of the dividend stream.

(d) Test the restriction $\beta = 1$ and interpret the result. If the restriction is satisfied re-estimate the present value model subject to this restriction and redo part (c).

4. **Fisher Hypothesis**

The data file contains United States quarterly data for the period September 1954 to December 2007 on the nominal interest rate, $r_t$, the price level, $p_t$, and inflation, $\pi_t$. The Fisher hypothesis states that $r_t$ fully reflects long-run movements in expected inflation, $E(\pi)$. 

\[ \text{fisher.*} \]
(a) To examine the Fisher hypothesis estimate the following linear regression

\[ r_t = \alpha + \beta \pi_t + u_t, \]

where actual inflation is used as a proxy for expected inflation and \( u_t \) is a disturbance term.

(b) Examine the properties of the estimated model by performing the following diagnostic tests.

i. Plot the ordinary least squares residuals and interpret their time series patterns.

ii. Test for autocorrelation of orders 1 to 4.

iii. Test for ARCH.

iv. Test for normality.

(c) If the Fisher hypothesis holds, \( \beta = 1 \). Test this restriction and interpret the result. If the restriction is satisfied re-estimate the present value model subject to this restriction and use the estimated model to generate an estimate of the real (ex post) interest rate \( i_t = r_t - \pi_t \).

5. Measuring Portfolio Performance

famafrench.*


(a) Estimate the Fama-French four factor model

\[ r_{it} - r_{ft} = \alpha + \beta_1 \text{EMKT}_t + \beta_2 \text{SMB}_t + \beta_3 \text{HML}_t + \beta_4 \text{MOM}_t + u_t, \]

for each of the 10 industry portfolios where EMKT\(_t\) is the excess return on the market, SMB\(_t\) is the size risk factor, HML\(_t\) is the value factor, MOM\(_t\) is the momentum factor and \( u_t \) is a disturbance term. Interpret the parameter estimates on all three risk factors.

(b) If \( \mu_p \) is the expected return on the portfolio, \( \mu_m \) is the expected return on the market, \( r_f \) is the risk-free rate, \( \sigma_p \) is the risk of a portfolio and \( \beta \) is the \( \beta \)-risk obtained from the single-factor CAPM, estimate the following portfolio performance measures for each of the
10 industry portfolios

Sharpe ratio:  \[ S = \frac{\mu_p - r_f}{\sigma_p} \]

Treynor index:  \[ T = \frac{\mu_p - r_f}{\beta} \]

Jensen’s alpha:  \[ \alpha = \mu_p - r_f - \beta(\mu_m - r_f). \]

Discuss the rankings of the portfolios based on each measure.

6. **Minimum Variance Portfolios**

The data are monthly observations for the period April 1990 to July 2004 on the equity prices of Exxon, General Electric, IBM and Microsoft and Walmart, together with the price of Gold, the S&P 500 index and a short-term interest rate.

(a) Consider the minimum variance portfolio regression model consisting of equity stocks in GE and Walmart

\[ r_{2t} = \alpha + \beta(r_{2t} - r_{1t}) + u_t, \]

where \( r_{1t} \) is the log return on GE, \( r_{2t} \) is the log return on Walmart and \( u_t \) is a disturbance term. Assume the mean and variance on the \( i^{th} \) asset are \( \mu_i \) and \( \sigma_i^2 \) with covariance \( \sigma_{12} \).

i. Show that the slope parameter satisfies

\[ \beta = \frac{\text{cov}(y_t, x_t)}{\text{var}(x_t)} = \frac{\sigma_2^2 - \sigma_{12}}{\sigma_1^2 + \sigma_2^2 - 2\sigma_{12}}, \]

which is the weight on GE in a minimum variance portfolio given by \( w_1 \).

ii. Letting \( w_1 \) and \( w_2 \) represent the weights attached to the two stocks in the portfolio, show that the intercept parameter satisfies

\[ \alpha = w_1 \mu_1 + (1 - w_1) \mu_2 = \mu_p, \]

which is the expected return on the minimum variance portfolio.

(b) Estimate the minimum variance portfolio regression model in part (a). Interpret the parameter estimates and the variance of the least squares residuals, without any degrees of freedom adjustment.

(c) Using the results in part (b)
i. Construct a test of an equal weighted portfolio, \( w_1 = w_2 = 0.5 \).

ii. Construct a test of portfolio diversification.

(d) Repeat parts (b) and (c) for the technology stocks, IBM and Microsoft.

(e) Repeat parts (b) and (c) for Exxon and Gold.

7. Microsoft and the Dot-Com Crisis

The data are monthly observations for the period April 1990 to July 2004 on the equity prices of Exxon, General Electric, IBM and Microsoft and Walmart, together with the price of Gold, the S&P 500 index and a short-term interest rate. The dot-com crash began on 10 March 2000 which led to very large falls in the equity value of Microsoft and technology stocks in general.

(a) Plot the price and log returns of Microsoft and identify the large movements in its share value during the period of the dot-com crisis.

(b) Estimate the CAPM model for Microsoft

\[
    r_{it} - r_{ft} = \beta_0 + \beta_1(r_{mt} - r_{ft}) + u_t,
\]

in which \( r_{ft} \) and \( r_{mt} \) are the risk free and market returns, respectively. Plot the residuals and perform the Jarque-Bera test of normality.

(c) To capture the effects of the dot-com crisis construct 11 indicator variables for each month of the crisis beginning with March 2000 and ending in January 2001

\[
    I_{1t} = \begin{cases} 
        1 : & \text{Mar. 2000} \\
        0 : & \text{Otherwise} 
    \end{cases}, \quad I_{2t} = \begin{cases} 
        1 : & \text{Apr. 2000} \\
        0 : & \text{Otherwise} 
    \end{cases}, \ldots,
\]

\[
    I_{11t} = \begin{cases} 
        1 : & \text{Jan. 2001} \\
        0 : & \text{Otherwise} 
    \end{cases}.
\]

The last date is chosen as this corresponds to the time of the positive correction in Microsoft shares. Now estimate the augmented CAPM regression

\[
    r_{it} - r_{ft} = \alpha + \beta_1(r_{mt} - r_{ft}) + \sum_{j=1}^{11} \beta_j I_{jt} + u_t.
\]
3.8. EXERCISES

i. Plot the residuals and redo the Jarque-Bera test of normality. Compare the result of this test with results of the test in part (b) based on the CAPM residuals.

ii. Interpret the parameter estimates of the dummy variables.

iii. Perform a joint test of the effects of the dot-com crisis on Microsoft by testing the restrictions $\beta_2 = \beta_3 = \cdots = \beta_{12}$.

8. The Retirement of Lee Raymond as the CEO of Exxon

The data are monthly observations on the equity price of EXXON together with the S&P 500 index and a short-term interest rate for the period January 1970 to March 2010. In December of 2005, Lee Raymond retired as the CEO of Exxon receiving the largest retirement package ever recorded of around $400m. How did the markets view this event?

(a) Estimate the market model for Exxon for the sub-period from January 1970 to September 2005

$$ r_t = \alpha + \beta_1 r_{mt} + u_t, $$

where $r_t$ is the log return on Exxon and $r_{mt}$ is the market log return computed from the S&P500 and $u_t$ is a disturbance term. Compute the abnormal returns over the 5-month event window from October 2005 to February 2006 by substituting in the actual values of $r_{mt}$ over this period into the estimated model.

(b) Estimate the augmented market model

$$ r_t = \alpha + \beta_1 r_{mt} + \sum_{j=1}^{5} \beta_{j+1} I_{jt} + u_t, $$

over the extended sample period from January 1970 to February 2006, where the 5 indicator variables are defined as

$$ I_{1t} = \begin{cases} 1 : \text{Oct. 2005} \\ 0 : \text{Otherwise} \end{cases}, \quad I_{2t} = \begin{cases} 1 : \text{Nov. 2005} \\ 0 : \text{Otherwise} \end{cases}, \quad \cdots, $$

$$ I_{5t} = \begin{cases} 1 : \text{Feb. 2001} \\ 0 : \text{Otherwise} \end{cases}. $$

Compare the abnormal returns over the event window with the estimates in part (a).

9. Properties of Ordinary Least Squares for Dependent Data
The minimum variance portfolio regression in Section 3.6 does not satisfy the assumption $E(x_t u_t) = 0$ given in (3.5) as the dependent variable $y_t$ is also present in the definition of the explanatory variable $x_t$. To investigate the sampling properties of the least squares estimator in this situation assume that the returns on two assets, $r_{1t}$ and $r_{2t}$, are summarised in terms of the following bivariate normal distribution

$$\begin{bmatrix} r_{1t} \\ r_{2t} \end{bmatrix} \sim iid N \left( \begin{bmatrix} 0.020877 \\ 0.013496 \end{bmatrix}, \begin{bmatrix} 0.011333 & 0.002380 \\ 0.002380 & 0.005759 \end{bmatrix} \right).$$

The minimum variance portfolio regression is defined as

$$y_t = \alpha + \beta x_t + u_t,$$

where $y_t = r_{2t}$ and $x_t = r_{2t} - r_{1t}$. From the properties of the minimum variance portfolio the true population parameter values are

$$\begin{align*}
\beta &= \frac{\sigma^2 - \sigma_{12}}{\sigma^2 - \sigma_{12} - 2\sigma_{12}} = \frac{0.005759 - 0.002380}{0.011333 + 0.005759 - 2 \times 0.002380} = 0.274 \\
\alpha &= w_1 \mu_1 + (1 - w_1) \mu_2 = 0.274 \times 0.020877 + 0.726 \times 0.013496 = 0.0155 \\
\sigma^2_u &= 0.004833 \\
\sigma^2_x &= 0.012331.
\end{align*}$$

(a) Using simulated data, estimate the slope coefficient $\beta$ in the minimum variance portfolio regression model for a sample of size $T = 20$. Repeat this estimation 1000 times, plot the histogram of the resultant estimates, and show that even in small samples the estimator $\hat{\beta}$ is well-centred.

(b) Demonstrate the efficiency of the least squares estimators of the parameters $\alpha$ and $\beta$ by allowing the sample to grow from $T = 10$ to $T = 10000$ in increments of 1 observation. For each sample size compute $\hat{\alpha}$ and $\hat{\beta}$, plot the resultant sequences of estimates and comment on the results.

(c) Compute and plot the asymptotic distribution of $(\hat{\beta} - \beta)/se(\hat{\beta})$ where the standard error is $se(\hat{\beta}) = \sigma^2_u/T \sigma^2_x$. Compare the asymptotic distribution with the simulated distribution of $\hat{\beta}$ constructed from samples of size $T = \{20, 50, 500, 1000\}$. 

Chapter 4

Dynamic Modelling with Stationary Variables

An important limiting feature of the linear regression model discussed in Chapter 3 is that all the variables that appear in the model are designated as being measured at the same point in time. To allow financial variables to adjust to shocks over time this model is now extended to incorporate dynamic effects. These effects enable the impact of shocks and transitions in a model to take place over a period of time rather than instantaneously. A primary class of dynamic models involves only a single variable and is called univariate or scalar. In a scalar dynamic model, a single dependent financial variable is explained using its own past history as well as lags of other relevant financial variables.

In financial applications, single dependent variable models are often extended to multivariate specifications in which several financial variables are jointly determined and modelled together. Such models are heavily used in central banks, treasuries, international agencies, and the financial industry. An important characteristic of the multivariate class of models investigated in this chapter is that each variable in the system is expressed as a linear function of its own lags as well as the lags of all of the other variables in the system. This type of model is known as a vector autoregression (VAR) model and was first explored systematically in a pioneering article by Mann and Wald (1943). The model has the distinguishing feature that each equation has the same set of explanatory variables, a feature which brings several advantages in practical implementation. Structural VAR models were also considered in the original article of Mann and Wald (1943). These models allow for contemporaneous interactive effects among the variables and have received much subsequent attention in recent years from researchers investigating the impact of policy changes on economic activity.
4.1 Stationarity

The models in this chapter, which use standard linear regression techniques, require that the variables involved satisfy a simplifying condition known as stationarity. Stationarity requires that all joint distributions of the time series remain unchanged when shifted over time. When this property is specialised to the first two moments of the distribution, a time series is said to be covariance stationary, so that the mean, variance, and autocovariances all remain invariant to the time periods in which they are calculated. These conditions fail when the mean, variance or autocovariances become time dependent. In particular, when the mean has a time trend or when the variance is time varying, then the time series has a nonstationary character. Such series are especially important in financial econometrics because many financial variables such as asset prices and interest rates have means and variances that exhibit changes over time. These topics form the subject matter of Chapters 5 and 6.

For the present a simple illustration will indicate the central idea. Consider Figures 4.1 and 4.2, which show the daily S&P 500 index and associated log returns, respectively. Assume that an observer takes a snapshot of the two series over two different decades (shown as lightly shaded regions in the figures) during this historical period: the first snapshot shows the trajectory of the series for the decade of the 1960s and the second shows the trajectory over 2000-2010. It is clear that the pattern of behaviour of the series in Figure 4.1 is completely different in these two decades. What the observer sees in 1960-1970 looks quite unlike what happens over 2000-2010.

The situation is different for the log returns plotted in Figure 4.2. Casual inspection suggests that the behaviour in the two shaded areas is remarkably similar given the long intervening time span of 30 years. Over both decades 1960-1970 and 2000-2010 the data fluctuate about the same level (which is approximately zero). The main difference is that during 2000-2010 the returns show evidence of greater variation largely because of the financial crisis of 2008-2009, with evidence of more extreme outliers than during 1960-1970. So there are differences between the periods even for returns. But the dominating difference occurs in levels, where the random wandering and fluctuating behavior of the S&P 500 distinguishes the 2000-2010 period from 1960-1970.

For present purposes it will be assumed that the series we deal with exhibit stationary behaviour over time, closer in form to that of the returns data shown in Figure 4.2 than the levels data shown in Figure 4.1. This assumption enhances the usefulness of historical data in estimating relationships, interpreting findings, and forecasting future behaviour by extrapolating from the past. In practical work, evidence for stationarity can be assessed using some of the techniques described in Chapter 5. Statistical corroboration of supporting assumptions like stationarity or specific forms of nonstationarity is important in empirical work because the validity of the methods used to make inferences often depends on these assumptions. The methods that are discussed in
4.1. STATIONARITY

Chapter 5 are helpful in this respect.

Figure 4.1: Snapshots of the time series of the S&P 500 index comprising monthly observations for the period February 1871 to September 2016.

Figure 4.2: Snapshots of the time series of returns to the S&P 500 computed from monthly observations for the period February 1871 to September 2016.
CHAPTER 4. DYNAMIC MODELLING WITH STATIONARY VARIABLES

4.2 Univariate Time Series Models

This section considers a class of dynamic models for a single dependent variable known as stationary time series models. The dynamics enter the model either through the lags of the dependent variable, or through the lags of the disturbances, or both. This type of model has been extensively studied and the discussion here will be brief and selective. Thorough theoretical treatments may be found in Brockwell and Davies (1991) or Hamilton (1994). A less theoretical treatment with more emphasis on implementation is in Martin, Hurn and Harris (2013).

4.2.1 Autoregressive Models

Specification

A linear dynamic model for the dependent variable, \( y_t \), in which movements in \( y_t \) are explained in terms of its own lags \( y_{t-j} \) with the longest lag included being the \( p^{th} \) lag, is given by

\[
y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + u_t, \quad u_t \sim iid \left(0, \sigma_u^2\right),
\]

(4.1)

in which \( \phi_0, \phi_1, \cdots, \phi_p \) are unknown parameters. Instead of assuming that the disturbance term is normally distributed, drawing on the discussion of the previous section concerning the importance of stationarity, the assumption is instead that \( u_t \) is independently and identically distributed, usually abbreviated as \( iid \). Equation (4.1) is referred to as an autoregressive model with \( p \) lags, or simply an AR(\( p \)). An AR(1) model is then given by

\[
y_t = \phi_0 + \phi_1 y_{t-1} + u_t, \quad u_t \sim iid \left(0, \sigma_u^2\right),
\]

(4.2)

where \( iid \left(0, \sigma_u^2\right) \) means \( iid \) with mean zero and variance \( \sigma_u^2 \). If the stability condition, \(|\phi_1| < 1\), is satisfied and the initial (start-up) condition of \( y_t \) is set to ensure a common mean and variance (which will be so if the initial condition is in the infinite past) then \( y_t \) is stationary. This condition is discussed in more detail in Chapter 5.

The unconditional mean and variance of an AR(1) model are easily derived. Applying the unconditional expectations operator to both sides of (4.2) gives

\[
E(y_t) = E(\phi_0 + \phi_1 y_{t-1} + u_t) = \phi_0 + \phi_1 E(y_{t-1}).
\]

One of the important implications of stationarity is that the unconditional expectations are the same at each point in time, so that \( E(y_t) = E(y_{t-1}) \). Applying this condition and rearranging gives

\[
E(y_t) = \frac{\phi_0}{1 - \phi_1}.
\]
4.2. UNIVARIATE TIME SERIES MODELS

The unconditional variance is defined as

\[ \text{var}(y_t) = \gamma_0 = E\{[y_t - E(y_t)]^2\}. \]

Recognising that

\[ y_t - E(y_t) = \phi_0 + \phi_1 y_{t-1} + u_t - \phi_0 - \phi_1 E(y_{t-1}) = \phi_1 [y_{t-1} - E(y_{t-1})] + u_t, \]

and then squaring both sides and taking unconditional expectations gives

\[
\begin{align*}
E\{[y_t - E(y_t)]^2\} &= \phi_1^2 E\{[y_{t-1} - E(y_{t-1})]^2\} + E(u_t^2) + 2 E\{[y_{t-1} - E(y_{t-1})]u_t\} \\
&= \phi_1^2 E\{[y_{t-1} - E(y_{t-1})]^2\} + E(u_t^2),
\end{align*}
\]

using the fact that \( E\{[y_{t-1} - E(y_{t-1})]u_t\} = 0 \). Moreover, because

\[ \gamma_0 = E\{[y_t - E(y_t)]^2\} = E\{[y_{t-1} - E(y_{t-1})]^2\}, \]

represents the variance of \( y_t \), it follows that

\[ \gamma_0 = \phi_1^2 \gamma_0 + \sigma_u^2, \]

and upon rearrangement

\[ \gamma_0 = \frac{\sigma_u^2}{1 - \phi_1^2}. \]

These results demonstrate that the unconditional mean and variance of a stationary AR(1) process are not functions of time.

**Estimation**

The unknown parameters \( \theta = \{\phi_0, \phi_1, \ldots, \phi_p, \sigma_u^2\} \) of the AR\( (p) \) model (4.1) are estimated by least squares. The residual sum of squares function for the AR\( (p) \) model is

\[
S = \sum_{t=p+1}^{T} u_t^2 = \sum_{t=p+1}^{T} (y_t - \hat{\phi}_0 - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2} - \cdots - \hat{\phi}_p y_{t-p})^2, \quad (4.3)
\]

where the sample sum of squares begins at \( t = p + 1 \) as \( p \) observations are lost because of the inclusion of \( p \) lags in the model. The effective sample size in estimation is therefore \( T - p \) rather than \( T \). This criterion function (4.3) is differentiated with respect to the parameters \( \{\phi_0, \phi_1, \ldots, \phi_p\} \) yielding \( p + 1 \) first-order conditions. Setting these derivatives to zero and rearranging gives the least squares estimators of the parameters. That is, estimation of the AR\( (p) \) model simply involves treating the lagged variables of \( y_t \) as regressors so estimation amounts to regressing \( y_t \) on a constant and its \( p \) lags.

After estimating the parameters the least squares residuals are computed as

\[
\hat{u}_t = y_t - \hat{\phi}_0 - \hat{\phi}_1 y_{t-1} - \hat{\phi}_2 y_{t-2} - \cdots - \hat{\phi}_p y_{t-p}, \quad (4.4)
\]
which are used to compute the residual variance
\[
\hat{\sigma}_u^2 = \frac{1}{T-p} \sum_{t=p+1}^{T} \hat{u}_t^2. \tag{4.5}
\]

### 4.2.2 Moving Average Models

#### Specification

An alternative way to introduce dynamics into univariate models is to allow for time dependence in the dependent variable \(y_t\) to be determined via the disturbance term \(u_t\). The specification of the model then has the form
\[
y_t = \psi_0 + u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \cdots + \psi_q u_{t-q}, \quad u_t \sim iid \left(0, \sigma_u^2\right), \tag{4.6}
\]
where \(\psi_0, \psi_1, \ldots, \psi_q\) are unknown parameters. Since the disturbance term is formulated as a moving weighted sum of current and past disturbances with \(\psi_i\) representing the weights, this model is referred to as a moving average model with \(q\) lags, or more simply as MA(\(q\)). An MA(1) model is then given by
\[
y_t = \psi_0 + u_t + \psi_1 u_{t-1}, \quad u_t \sim iid \left(0, \sigma_u^2\right), \tag{4.7}
\]
where the condition \(|\psi_1| \leq 1\) is often imposed.

Applying the unconditional expectations operator to both sides of equation (4.7) gives the unconditional mean of the MA(1) process
\[
E(y_t) = E(\psi_0 + u_t + \psi_1 u_{t-1}) = \psi_0 + E(u_t) + \psi_1 E(u_{t-1}) = \psi_0,
\]
where the property \(E(u_t) = E(u_{t-1}) = 0\) has been used. Using the result that \(E(u_t u_{t-1}) = 0\) by virtue of the iid assumption on \(u_t\), the unconditional variance is
\[
\gamma_0 = E\{[y_t - E(y_t)]^2\} = E[(u_t + \psi_1 u_{t-1})^2] = \sigma_u^2 (1 + \psi_1^2).
\]
Just as in the case of the AR(1) model, these results demonstrate that the unconditional mean and variance of a stationary MA(1) process are not functions of time.

#### Estimation

Unlike the AR(\(p\)) model where estimation of the parameters is based on ordinary least squares, the unknown parameters \(\theta = \{\psi_0, \psi_1, \ldots, \psi_q, \sigma_u^2\}\) of the MA(\(q\)) model in (4.6) are estimated by nonlinear methods. The residual sum of squares function for the MA(\(q\)) model is
\[
S = \sum_{t=q+1}^{T} u_t^2 = \sum_{t=q+1}^{T} (y_t - \psi_0 - \psi_1 u_{t-1} - \psi_2 u_{t-2} - \cdots - \psi_q u_{t-q})^2, \tag{4.8}
\]
where the sample sum of squares begins at $t = q+1$ as $q$ observations are lost because the disturbances $u_q, u_{q-1}, \cdots, u_1$ are (unobserved) initializations. The effective sample size for estimating the parameters is therefore $T - q$.

The residual sum of squares function is differentiated with respect to the parameters $\{\psi_0, \psi_1, \cdots, \psi_p\}$ yielding $q+1$ first-order conditions. Numerical complications in estimating the MA($q$) model’s parameters arise in this step as the unknown parameters appear directly in (4.8) as well as indirectly through the lagged disturbance terms $u_{t-1}$ to $u_{t-q}$. For example, the derivative with respect to the parameter $\psi_i$ is

$$\frac{\partial S}{\partial \psi_i} = 2 \sum_{t=q+1}^{T} u_t \left( -u_{t-i} - \psi_1 \frac{\partial u_{t-1}}{\partial \psi_i} - \psi_2 \frac{\partial u_{t-2}}{\partial \psi_i} - \cdots - \psi_q \frac{\partial u_{t-q}}{\partial \psi_i} \right).$$

The existence of this indirect channel means that it is now no longer possible to rearrange the first-order conditions to derive direct analytical solutions of the unknown parameters as was the case with the AR model. Even though no analytical solution is available to compute the parameter estimates a numerical solution is nonetheless available by using an iterative algorithm. A general introduction to these algorithms is given in Appendix D. An alternative estimation strategy is developed in the exercises which uses a two-step procedure that does not require an iterative solution.

Having estimated the parameters by numerical methods the nonlinear least squares residuals are computed recursively as follows

$$\hat{u}_t = y_t - \hat{\psi}_0 - \hat{\psi}_1 \hat{u}_{t-1} - \hat{\psi}_2 \hat{u}_{t-2} - \cdots - \hat{\psi}_p \hat{u}_{t-q}, \quad (4.9)$$

and these residuals may be used to compute the residual variance, $\hat{\sigma}^2_u$, as in (4.5), with $p$ replaced by $q$ and with initial conditions $u_q = \cdots = u_1 = 0$ set to the mean value $E(u_t) = 0$.

### 4.2.3 Autoregressive-Moving Average Models

**Specification**

Autoregressive and moving average forms may be combined to yield a model containing both components called an autoregressive-moving average (ARMA) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + u_t + \psi_1 u_{t-1} + \cdots + \psi_q u_{t-q}, \quad u_t \sim iid \left(0, \sigma^2_u\right).$$

This model has $p$ autoregressive lags and $q$ moving average lags and is therefore denoted as ARMA($p, q$). In the case of an ARMA(1,1) model

$$y_t = \phi_0 + \phi_1 y_{t-1} + u_t + \psi_1 u_{t-1}, \quad u_t \sim iid \left(0, \sigma^2_u\right), \quad (4.10)$$
the unconditional mean and variance are, respectively,

\[ E(y_t) = \frac{\phi_0}{1 - \phi_1}, \quad \gamma_0 = \frac{1 + \psi_1^2 + 2\phi_1\psi_1}{1 - \phi_1^2} \sigma_u^2. \]

As with the AR and MA models, the first two moments of the ARMA model are not functions of time.

Estimation

In view of its MA component, ARMA models also require numerical methods such as nonlinear or iterative least squares procedures to estimate the unknown parameters. In the case of the ARMA(1,1) model the residual sum of squares function is

\[ S = \sum_{t=2}^{T} u_t^2 = \sum_{t=2}^{T} (y_t - \phi_0 - \phi_1 y_{t-1} - \psi_1 u_{t-1})^2, \quad (4.11) \]

where the sample begins at \( t = 2 \) as a result of the presence of the lagged dependent variable \( y_{t-1} \) and the lagged disturbance \( u_{t-1} \). Differentiating with respect to the unknown parameters gives

\[
\frac{\partial S}{\partial \phi_0} = 2 \sum_{t=2}^{T} u_t (-1 - \psi_1 \frac{\partial u_{t-1}}{\partial \phi_0}) \\
\frac{\partial S}{\partial \phi_1} = 2 \sum_{t=2}^{T} u_t (-y_{t-1} - \psi_1 \frac{\partial u_{t-1}}{\partial \phi_1}) \\
\frac{\partial S}{\partial \psi_1} = 2 \sum_{t=2}^{T} u_t (-u_{t-1} - \psi_1 \frac{\partial u_{t-1}}{\partial \psi_1}).
\]

This is a nonlinear set of equations in the unknown parameters which requires an iterative solution. As with the AR(\( p \)) and MA(\( q \)) models once the parameter estimates are obtained the least squares residuals \( \hat{u}_t \) are computed as well as the residual variance, \( \hat{\sigma}_u^2 \), based on (4.5) with the appropriate correction for lost observations.

4.2.4 Regression Models with Dynamics

The regression models discussed in Chapter 3 have the property that dependent and explanatory variables are all measured concurrently at the same time \( t \). To allow for dynamic effects, regression models may be combined
with the ARAMA class of models. Some examples are as follows.

\begin{align*}
  y_t &= \alpha + \beta x_t + u_t, \quad u_t = \phi_1 u_{t-1} + v_t \quad \text{[AR disturbance]} \\
  y_t &= \alpha + \beta x_t + u_t, \quad u_t = v_t + \psi_1 v_{t-1} \quad \text{[MA disturbance]} \\
  y_t &= \alpha + \beta x_t + \lambda y_{t-1} + u_t \quad \text{[Lagged dependent]} \\
  y_t &= \alpha + \beta x_t + \gamma x_{t-1} + u_t \quad \text{[Lagged explanatory]} \\
  y_t &= \alpha + \beta x_t + \lambda y_{t-1} + \gamma x_{t-1} + u_t \quad \text{[Joint specification]} \\
  y_t &= \alpha + \beta x_t + \lambda y_{t-1} + \gamma x_{t-1} + u_t \quad \text{[Joint specification]} \\
  u_t &= \phi_1 u_{t-1} + v_t + \psi_1 v_{t-1}.
\end{align*}

A natural mechanism for the introduction of dynamics in linear regression occurs in the case of models of forward market efficiency. Lags arise in these models for two reasons. First, the forward rate acts as a natural predictor of future spot rates. Second, if the data are overlapping so that the maturity of the forward rate is longer than the frequency of observations, the disturbance term has an induced moving average structure. This point is taken up in Exercise 2.

One reason for including dynamics in a regression model is to correct for potential misspecification problems that arise from incorrectly excluding explanatory variables. In Chapter 3, it was shown how misspecification of this type may be detected using the LM autocorrelation test applied to the residuals of the estimated regression model.

## 4.3 Autocorrelation and Partial Autocorrelations

In addition to the unconditional mean and variance not being functions of time, the discussion on stationarity in Section 4.1 emphasises the autocovariance of a stationary process is also invariant with respect to the time period in which they are computed.

The autocovariance at order \( k \) gives the covariance between \( y_t \) and \( y_{t-k} \). The first \( k \) autocovariances of \( y_t \) are therefore defined as

\[
\gamma_1 = \text{E}\{ [y_t - \text{E}(y_t)] [y_{t-1} - \text{E}(y_{t-1})] \} \\
\gamma_2 = \text{E}\{ [y_t - \text{E}(y_t)] [y_{t-2} - \text{E}(y_{t-2})] \} \\
\vdots \quad \vdots \quad \vdots \\
\gamma_k = \text{E}\{ [y_t - \text{E}(y_t)] [y_{t-k} - \text{E}(y_{t-k})] \}. 
\]

The autocorrelation function is obtained by standardising the autocovariances by the variance, \( \gamma_0 \), and is given by

\[ \rho_k = \frac{\gamma_k}{\gamma_0}. \] (4.12)

Consider again the AR(1) model in (4.2). The first-order autocovariance is

\[
\gamma_1 = \text{E}\{ [\phi_1 y_{t-1} - \phi_1 \text{E}(y_{t-1}) + u_t] [y_{t-1} - \text{E}(y_{t-1})] \} 
\]
where the second step is based on the result, stated previously, that $E\{[y_t - E(y_{t-1})]u_t\} = 0$. It follows that the $k^{th}$ autocovariance for $k > 0$ is

$$\gamma_k = \phi_1^k \gamma_0. \quad (4.13)$$

By virtue of the definition $\gamma_k = \gamma_{-k}$, for general integer $k$ we have $\gamma_k = \phi_1^{|k|} \gamma_0$. From this result it can be seen that the autocovariances of a stationary AR model are not functions of time but do depend on the lag parameter $k$ in the autocovariance.

The autocorrelation function (ACF) of the AR(1) model is

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^{|k|}. \quad (4.14)$$

For $0 < \phi_1 < 1$, the autocorrelation function of $y_t$ declines exponentially for increasing $k$ so that the effects of previous values on $y_t$ gradually diminish. For $-1 < \phi_1 < 0$, the autocorrelation function of $y_t$ alternates in sign as $k$ is even or odd and its modulus declines exponentially. For higher order AR models the properties of the ACF are more complicated in analytic form but nonetheless follow similar patterns with their moduli having exponential decay with increasing $k$ whenever the AR model is stable.

The pattern of autocovariances and associated autocorrelations for the MA(1) model in 4.7 is quite different to the AR(1) case. The first-order autocovariance is

$$\gamma_1 = E\{[y_t - E(y_t)][y_{t-1} - E(y_{t-1})]\} = E\{(u_t + \psi_1 u_{t-1})(u_{t-1} + \psi_1 u_{t-2})\} = \psi_1 \sigma_u^2,$$

which uses the properties

$$E(u_t u_{t-1}) = E(u_{t-1} u_{t-2}) = E(u_t u_{t-2}) = 0, \quad E(u_t^2) = E(u_{t-1}^2) = \sigma_u^2.$$

For autocovariances of $y_t$ for $k > 1$, $\gamma_k = 0$. The ACF of an MA(1) model is summarised as

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \begin{cases} \frac{\psi_1}{1 + \psi_1^2} & : \quad k = 1 \\ 0 & : \quad \text{otherwise.} \end{cases} \quad (4.15)$$

This result contrasts with the ACF of the AR(1) model as there is now a single spike in the ACF at lag 1 which corresponds to the lag length of the model. In a similar way, the ACF of an MA($q$) model has non-zero values for the first $q$ lags and zero thereafter.
The first-order autocorrelation of the ARMA(1,1) model in equation (4.10) is
\[ \rho_1 = \frac{(1 + \phi_1 \psi_1)(\phi_1 + \psi_1)}{1 + \psi_1^2 + 2\phi_1 \psi_1}, \]
while higher-order autocorrelations are given by
\[ \rho_k = \phi_1 \rho_{k-1}, \quad k \geq 2. \]
Setting \( \psi_1 = 0 \) produces the autocorrelation function of the AR(1) model given in (4.2). Similarly, setting \( \phi_1 = 0 \) results in the autocorrelation function of the MA(1) model given in (4.15). For the special case where \( \psi_1 = -\phi_1 \), the AR and MA components cancel and the model contains no dynamics since it reduces to
\[ y_t = \frac{\phi_0}{1 - \phi_1} + \epsilon_t, \quad \Rightarrow y_t \sim iid \left( \frac{\phi_0}{1 - \phi_1}, \sigma_u^2 \right), \]
with \( \rho_k = 0 \) for all \( k \geq 1. \)

To compute the sample ACF, the following sequence of AR models may be estimated equation-by-equation by ordinary least squares
\[
\begin{align*}
y_t &= \phi_{10} + \rho_1 y_{t-1} + u_{1t}, \\
y_t &= \phi_{20} + \rho_2 y_{t-2} + u_{2t}, \\
&\vdots \quad \vdots \\
y_t &= \phi_{k0} + \rho_k y_{t-k} + u_{kt},
\end{align*}
\]
giving the estimated ACF \( \{\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_k\} \). The notation adopted for the constant term in the above regressions emphasises that this term differs for each equation.

To illustrate, consider United States monthly data from February 1871 to September 2016 on equity returns\(^1\) expressed as a percentage, given by
\[ re_t = 100 \times (\log P_t - \log P_{t-1}), \quad (4.16) \]
where \( P_t \) is the equity price index. The ACF and PACF of the equity returns are computed by means of a sequence of regressions. Specifically, the ACF for lags 1 to 3 is computed using the following three regressions (standard errors in parentheses):
\[
\begin{align*}
re_t &= 0.253 + 0.284 re_{t-1} + \hat{u}_{1t}, \\
(0.094) & \quad (0.023) \\
re_t &= 0.350 + 0.005 re_{t-2} + \hat{u}_{2t}, \\
(0.098) & \quad (0.024)
\end{align*}
\]
\(^{1}\)Strictly speaking these are log equity returns expressed as percentages, but the “log” will be dropped for expositional ease when referring to \( re_t \) as defined in equation (4.16). The same applies to the log dividend returns, \( rd_t \) which are formally defined in equation (4.17).
\[ re_t = 0.365 - 0.041 \, re_{t-3} + \hat{u}_{3t}. \]

The estimated ACF is \( \hat{\rho}_1 = 0.284, \hat{\rho}_2 = 0.005, \hat{\rho}_3 = -0.041 \). Notice that the standard errors on each of the \( \hat{\rho}_k \) are approximately the same. In fact a useful result is that the standard errors of the autocorrelations are approximately \( 1/\sqrt{T} \), where \( T \) is the sample size. These standard errors are commonly referred to as Bartlett standard errors (see, for example, Brockwell and Davis, 1991).

Another measure of the dynamic properties of AR models is the partial autocorrelation function (PACF) at lag \( k \), which measures the relationship between \( y_t \) and \( y_{t-k} \) but now with the intermediate lags included in the regression model, so that their effects are controlled for. The PACF at lag \( k \) is denoted by \( \phi_{kk} \). By implication the PACF for an AR\((p)\) model is zero for lags greater than \( p \). For example, in the AR(1) model the PACF has a spike at lag 1 and thereafter is \( \phi_{kk} = 0, \forall \ k > 1 \). This is in contrast to the ACF which in general has non-zero values for higher lags, as seen in the simple AR(1) model above. Note that by construction the ACF and PACF at lag 1 are equal to each other.

To compute the sample PACF the following AR models are estimated equation-by-equation by ordinary least squares

\[
\begin{align*}
y_t &= \phi_{10} + \phi_{11}y_{t-1} + u_t, \\
y_t &= \phi_{20} + \phi_{21}y_{t-1} + \phi_{22}y_{t-2} + u_{2t}, \\
y_t &= \phi_{30} + \phi_{31}y_{t-1} + \phi_{32}y_{t-2} + \phi_{33}y_{t-3} + u_{3t} \\
&\quad \vdots \\
y_t &= \phi_{k0} + \phi_{k1}y_{t-1} + \phi_{k2}y_{t-2} + \cdots + \phi_{kk}y_{t-k} + u_{kt},
\end{align*}
\]

where the estimated PACF is therefore given by \( \{\hat{\phi}_{11}, \hat{\phi}_{22}, \cdots, \hat{\phi}_{kk}\} \).

The PACF for lags 1 to 3 for the equity returns data is then obtained from

\[
\begin{align*}
re_t &= 0.253 + 0.284 \, re_{t-1} + \hat{u}_{1t}, \\
re_t &= 0.272 + 0.307 \, re_{t-1} - 0.082 \, re_{t-2} + \hat{u}_{2t}, \\
re_t &= 0.277 + 0.305 \, re_{t-1} - 0.075 \, re_{t-2} - 0.021 \, re_{t-3} + \hat{u}_{3t}.
\end{align*}
\]

The estimated sample PACF is \( \{\hat{\phi}_{11} = 0.284, \hat{\phi}_{22} = -0.082, \hat{\phi}_{33} = -0.021\} \).

Plots of the sample autocorrelation and partial autocorrelation functions can yield valuable insights into the appropriate dynamic model for a process. Consider United States monthly data from February 1871 to September 2016 on dividend returns expressed as a percentage,

\[
rd_t = 100 \times (\log D_t - \log D_{t-1}), \quad (4.17)
\]
4.4. MEAN AVERSION AND REVERSION IN RETURNS

A well-known empirical result in finance is that asset returns exhibit mean aversion (positive autocorrelation) for relatively short time horizons and mean reversion (negative autocorrelation) for longer time horizons (Fama and French, 1988; Kim, Nelson and Startz, 1991). This property is illustrated by computing the autocorrelations of log returns on the NASDAQ share index for alternative frequencies. Using monthly, quarterly and annual frequencies for the period 1989 to 2009, the following results are obtained from estimating a simple model.
AR(1) model in each case:

- **Monthly**: \( r_t = 0.599 + 0.131 r_{t-1} + \hat{u}_{1t} \),
- **Quarterly**: \( r_t = 1.950 + 0.058 r_{t-1} + \hat{u}_{2t} \),
- **Annual**: \( r_t = 8.974 − 0.131 r_{t-1} + \hat{u}_{3t} \),

where the \( \hat{u}_{it} \) are the ordinary least squares residuals. There appears to be mean aversion in returns for time horizons less than a year as the first-order autocorrelation is positive for monthly and quarterly returns. By contrast, there is mean reversion for horizons of at least a year as the first-order autocorrelation is now negative with a value of \(-0.131\) for annual returns.

To understand the change in the autocorrelation properties of returns over different maturities, consider the following model proposed by Poterba and Summers (1988),

\[
p_t = f_t + u_t,
\]
\[
f_t = f_{t-1} + v_t,
\]
\[
u_t = \phi_1 u_{t-1} + w_t,
\]

where \( p_t \) is the log of prices, \( f_t \) is the log of market fundamentals and \( v_t \) and \( u_t \) are disturbance terms assumed to be independent of each other. The disturbance term \( u_t \) represents transient deviations of \( p_t \) from \( f_t \), while \( v_t \) captures shocks to the market fundamentals.

The 1-period log return on the asset is

\[
r_{t} = p_t - p_{t-1} = v_t + u_t - u_{t-1},
\]

and the \( k \)-period return is

\[
r_{t}(k) = p_t - p_{t-k} = r_{t} + r_{t-1} + \cdots + r_{t-k+1}
\]
\[
= (v_t + u_t - u_{t-1}) + (v_{t-1} + u_{t-1} - u_{t-2}) + \cdots
\]
\[
+ (v_{t-k+1} + u_{t-k+1} - u_{t-k})
\]
\[
= v_t + v_{t-1} + \cdots + v_{t-k+1} + u_t - u_{t-k}.
\]

Using the property that \( E(r_{t}) = 0 \), the autocovariance is

\[
\gamma_k = E[(p_t - p_{t-k})(p_{t-k} - p_{t-2k})]
\]
\[
= E[(v_t + v_{t-1} + \cdots + v_{t-k+1} + u_t - u_{t-k})
\]
\[
\times (v_{t-k} + v_{t-k-1} + \cdots + v_{t-2k+1} + u_{t-k} - u_{t-2k})]
\]
\[
= E(u_t u_{t-k}) - E(u_t u_{t-2k}) - E(u_{t-k}^2) + E(u_{t-k} u_{t-2k})
\]
\[
= 2 E(u_t u_{t-k}) - E(u_t u_{t-2k}) - E(u_{t-k}^2).
\]

Since \( u_t \) is an AR(1) process, it follows that

\[
\gamma_k = \frac{\sigma_w^2}{1 - \phi_1^2} (2\phi_1^k - \phi_2^{2k} - 1).
\]
For small values of $k$, $\gamma_k$ can be positive, but in the limit as $k \to \infty$, $\gamma_k$ eventually becomes negative, because

$$\lim_{k \to \infty} \gamma_k = -\frac{\sigma_w^2}{1 - \phi_1^2} = -\sigma_u^2.$$

### 4.5 Vector Autoregressive Models

Once a decision is made to move into a multivariate setting, it becomes difficult to delimit a single variable as the dependent variable to be explained in terms of all the other variables. Given the interdependence of economic systems and the joint decision making involved in portfolio choice, it is reasonable to suppose that variables are jointly determined within a system of equations. This way of thinking about economic data has a long history that goes back to the earliest empirical work on price determination by demand and supply. The approach became firmly established during the 1940s with the work of the Cowles Commission researchers at the University of Chicago. From a time series perspective, Mann and Wald (1943) made the fundamental contribution of developing a complete theory of estimation and inference for systems of equations in vector autoregressive (VAR) form under stationarity conditions. Their work also allowed for contemporaneous simultaneous equations effects within the VAR framework, corresponding to what is now known as structural vector autoregressions (SVARs).

The adoption of a VAR modelling framework in empirical finance has many advantages, which partly explains its great popularity in applied research. Among these advantages are the following.

(i) Estimation is straightforward, involving the application of ordinary least squares to each equation in the VAR.

(ii) The VAR system provides a convenient framework to forecast financial variables.

(iii) The model provides a basis for performing causality tests between financial variables.

(iv) The dynamics of the VAR can be modelled using impulse response analysis, which reveal the effects of shocks on the system variables.

(v) The volatility of financial variables can be decomposed in terms of their risk components.

(vi) Theoretical models in finance can be tested through the imposition of restrictions on the VAR parameters.
There is a convenient direct link between a VAR and an important type of model which will be introduced in Chapter 6, known as an error correction model.

### 4.5.1 Specification

A major empirical contribution within this approach to modelling, complete with some new methodology for studying a system’s impulse responses to shocks, was provided by Sims (1980) using United States data on nominal interest rates, money, prices and output. The approach proposed to treat all variables as endogenous by specifying a model in which each of the variables has an equation that explains its movements using past information on all the variables in the system. An important distinguishing feature of this approach is that each equation has precisely the same set of explanatory variables. This type of model is known as an unrestricted vector autoregression or simply a VAR.

An example of a trivariate VAR\((p)\) in this framework is

\[
y_{1t} = \phi_{10} + \sum_{i=1}^{p} \phi_{11,i} y_{1t-i} + \sum_{i=1}^{p} \phi_{12,i} y_{2t-i} + \sum_{i=1}^{p} \phi_{13,i} y_{3t-i} + u_{1t},
\]

\[
y_{2t} = \phi_{20} + \sum_{i=1}^{p} \phi_{21,i} y_{1t-i} + \sum_{i=1}^{p} \phi_{22,i} y_{2t-i} + \sum_{i=1}^{p} \phi_{23,i} y_{3t-i} + u_{2t}, \tag{4.18}
\]

\[
y_{3t} = \phi_{30} + \sum_{i=1}^{p} \phi_{31,i} y_{1t-i} + \sum_{i=1}^{p} \phi_{32,i} y_{2t-i} + \sum_{i=1}^{p} \phi_{33,i} y_{3t-i} + u_{3t},
\]

where \(y_{1t}, y_{2t}\) and \(y_{3t}\) are the jointly dependent variables, \(p\) is a prescribed lag length which is the same for all equations and \(u_{1t}, u_{2t}, u_{3t}\) are disturbance terms. Even though this model contains just three variables, the total number of parameters can become quite large especially if long lag structures are entertained. In the case of the trivariate VAR the number of unknown parameters in each equation is \(1 + 3p\) making the total number of parameters for the whole model \(3(1 + 3p)\). If \(p = 6\), for instance, then the total number of unknown parameters is \(3(1 + 3 \times 6) = 57\).

Higher dimensional VARs containing \(N\) variables \(\{y_{1t}, y_{2t}, \ldots, y_{Nt}\}\), are specified in the same way as they are for the trivariate VAR in (4.18). In matrix notation the VAR is conveniently represented as

\[
y_t = \Phi_0 + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \cdots + \Phi_p y_{t-p} + u_t, \tag{4.19}
\]
where the parameter matrices are given by
\[
\Phi_0 = \begin{bmatrix} \phi_{10} \\ \phi_{20} \\ \vdots \\ \phi_{K0} \end{bmatrix}, \quad \Phi_i = \begin{bmatrix} \phi_{11,i} & \phi_{12,i} & \cdots & \phi_{1N,i} \\ \phi_{21,i} & \phi_{22,i} & \cdots & \phi_{2N,i} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{N1,i} & \phi_{N2,i} & \cdots & \phi_{NN,i} \end{bmatrix}.
\]

The disturbances \( u_t = [u_{1t}, u_{2t}, \ldots, u_{Nt}]' \sim iid (0, \Omega) \) are independent over \( t \) with zero mean and covariance matrix
\[
\Omega = E(u_t u_t') = \begin{bmatrix} \sigma^2_1 & \sigma_{12} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma^2_2 & \cdots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1} & \sigma_{N2} & \cdots & \sigma^2_N \end{bmatrix}. \tag{4.20}
\]

This matrix has two properties. First, it is a symmetric matrix so that the upper triangular part of the matrix mirrors the lower triangular part
\[
\sigma_{ij} = \sigma_{ji}, \quad i \neq j.
\]

Second, the disturbance terms in each equation are generally correlated with the disturbances of other equations, so that
\[
\sigma_{ij} \neq 0, \quad i \neq j.
\]

This last property allows for interdependence amongst the shocks that drive the VAR system. It is particularly relevant when undertaking impulse response analysis and in computing variance decompositions, topics that are addressed at a later stage.

### 4.5.2 Estimation

Despite the VAR in (4.19) being a multivariate system of equations with lagged values of each variable potentially influencing all the others, estimation of the parameters \( \{\Phi_0, \Phi_1, \Phi_2, \cdots, \Phi_p\} \) is performed by simply applying ordinary least squares to each equation one at a time. This strategy is appropriate because the set of explanatory variables is the same in each equation and there are no restrictions on the coefficients of the system.

To estimate the covariance matrix \( \Omega \) in (4.20) let \( \hat{u}_t = \{\hat{u}_{1t}, \hat{u}_{2t}, \cdots, \hat{u}_{Nt}\} \) represent the least squares residuals for each equation in the VAR. An estimate of the covariance matrix is computed using the sample residual moment matrix defined by
\[
\hat{\Omega} = \begin{bmatrix} \hat{\sigma}^2_1 & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1N} \\ \hat{\sigma}_{21} & \hat{\sigma}^2_2 & \cdots & \hat{\sigma}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{N1} & \hat{\sigma}_{N2} & \cdots & \hat{\sigma}^2_N \end{bmatrix} = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t \hat{u}_t', \tag{4.21}
\]
or more explicitly,

$$
\hat{\Omega} = \frac{1}{T} \begin{bmatrix}
\sum_{t=1}^{T} \hat{u}_{1t}^2 & \sum_{t=1}^{T} \hat{u}_{1t}\hat{u}_{2t} & \cdots & \sum_{t=1}^{T} \hat{u}_{1t}\hat{u}_{Nt} \\
\sum_{t=1}^{T} \hat{u}_{2t}\hat{u}_{1t} & \sum_{t=1}^{T} \hat{u}_{2t}^2 & \cdots & \sum_{t=1}^{T} \hat{u}_{2t}\hat{u}_{Nt} \\
\vdots & \vdots & \ddots & \vdots \\
\sum_{t=1}^{T} \hat{u}_{Nt}\hat{u}_{1t} & \sum_{t=1}^{T} \hat{u}_{Nt}\hat{u}_{2t} & \cdots & \sum_{t=1}^{T} \hat{u}_{Nt}^2
\end{bmatrix}.
$$

As an example, consider a bivariate model for equity returns as defined in equation (4.16), \( r_{et} \), by adding as an explanatory variable lagged dividend returns as defined in equation (4.17), \( r_{dt} \). Such an extension is justified on theory grounds given the link between equity prices and dividends that is established in the present value model and discussed in Chapter 2. Equally important is the need for a model to explain dividend returns. A natural specification is to include as explanatory variables both own lags and lags of equity returns in this model. Then, treating equity returns, \( r_{et} \), and dividend returns, \( r_{dt} \), as potentially endogenous and jointly determined. Consequently, a VAR(6) model is estimated with monthly United States data from 1871 to 2016. The parameter estimates with standard errors in parentheses, are given in Table 4.1.

Notice that the parameter estimates of the effects of dividend returns on equity returns are borderline significant for lags 2 and 6 at the 5% level. The parameter estimates of the effects of equity returns on dividend returns at lags 2, 3, 5 and 6 are also statistically significant at this level, so that lagged values of \( r_{et} \) are important in explaining the behaviour of \( r_{dt} \).

### Table 4.1

Parameter estimates of a bivariate VAR(6) model for United States monthly equity and dividend returns for the period February 1871 to September 2016.

<table>
<thead>
<tr>
<th>Lag</th>
<th>Equity Returns</th>
<th>Dividend Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( r_{et} )</td>
<td>( r_{dt} )</td>
</tr>
<tr>
<td></td>
<td>( r_{et} )</td>
<td>( r_{dt} )</td>
</tr>
<tr>
<td>1</td>
<td>0.297 (0.024)</td>
<td>-0.049 (0.187)</td>
</tr>
<tr>
<td>2</td>
<td>-0.070 (0.025)</td>
<td>0.519 (0.254)</td>
</tr>
<tr>
<td>3</td>
<td>-0.029 (0.025)</td>
<td>-0.248 (0.250)</td>
</tr>
<tr>
<td>4</td>
<td>0.030 (0.025)</td>
<td>0.318 (0.251)</td>
</tr>
<tr>
<td>5</td>
<td>0.052 (0.025)</td>
<td>-0.231 (0.254)</td>
</tr>
<tr>
<td>6</td>
<td>-0.005 (0.024)</td>
<td>-0.341 (0.186)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.264 (0.097)</td>
<td>0.017 (0.012)</td>
</tr>
</tbody>
</table>
4.5. VECTOR AUTOREGRESSIVE MODELS

The estimated covariance matrix of the residuals in equation (4.20) is constructed using the parameter estimates reported in Table 4.1 giving

\[
\hat{\Omega} = \begin{bmatrix}
14.886 & -0.092 \\
-0.092 & 0.241
\end{bmatrix}.
\] (4.22)

This estimate shows that there is a large difference in the residual variance in the two equations and that there is a negative covariance between the residuals of the equity and dividend return equations. As discussed next, the matrix \( \hat{\Omega} \) provides an important input concerning model fit, one that is particularly useful in methods designed to select lag length specification in the VAR.

4.5.3 Lag Length Selection

An important part of the specification of a VAR model is the choice of the lag structure whose key parameter is the lag length order \( p \). If the lag length is too short, there is a risk that aspects of the dynamic mechanism are excluded from the model. If the lag structure is too long then there are redundant lags which can reduce the precision of the parameter estimates, thereby raising the standard errors and yielding \( t \) statistics that may be biased downwards.

In choosing the lag structure of a VAR, care must be exercised in relation to the sample size as degrees of freedom quickly diminish for even moderate lag lengths. For each integer increase in the lag length, an additional matrix of coefficients must be estimated. In a \( K \) dimensional system this means an additional \( K^2 \) coefficient parameters for each extra lag.

For these reasons, an important practical consideration in constructing and estimating a VAR(\( p \)) model is the choice of the lag order \( p \). A common data-driven approach to selecting lag order is to use information criteria. These criteria are scalar statistics that provide a simple but effective way of balancing improvements in sample period fit of the equations against the loss of degrees of freedom in estimation that results from increasing the lag order. Many such criteria are now available for use in econometric work.

The three most commonly used information criteria (IC) for selecting a parsimonious time series model are the Akaike information criterion (AIC) (Akaike, 1974, 1976), the Hannan information criterion (HIC) (Hannan and Quinn, 1979; Hannan, 1980) and the Schwarz information criterion (SIC) (Schwarz, 1978). The versions reported here follow Lütkepohl (2005). If \( K \) is the number of variables in the VAR(\( p \)) system, then criteria as follows

\[
AIC = \log |\hat{\Omega}| + \frac{2pK^2}{T},
\] (4.23)

\[
HIC = \log |\hat{\Omega}| + \frac{2\log(\log(T))}{T}pK^2,
\] (4.24)
In these expressions, $\hat{\Omega}$ is an estimate of the covariance matrix given in equation (4.21) with the numerical estimate reported in (4.22) for the example above. In the scalar case, the determinant of the estimated covariance matrix, $|\hat{\Omega}|$, is replaced by the estimated residual variance, $\hat{\sigma}^2_u$.

Choosing an IC optimal lag order using any of the above criteria requires the following steps.

**Step 1:** Choose a maximum number of lags, $p_{\text{max}}$, for the VAR model. This choice may be informed by the ACFs and PACFs of the data, the frequency with which the data are observed and the sample size.

**Step 2:** Estimate the model sequentially for all lags up to and including $p_{\text{max}}$. For each regression, compute the relevant information criterion, holding the sample size fixed.

**Step 3:** Choose the specification corresponding to the minimum values of the information criterion. In some cases there will be disagreement between different information criteria on the choice of lag length. The final decision is then a matter of individual judgement.

The bivariate VAR(6) for equity and dividend returns in Table 4.1 used an arbitrarily chosen maximum lag length $p_{\text{max}} = 12$. In order to provide a data-determined choice, the information criteria outlined above in 4.5.3 can be used. For example, calculations of the AIC, HIC, and SIC values for this VAR for lags from 1 to 8 are as follows:

<table>
<thead>
<tr>
<th>Lag:</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIC</td>
<td>7.080</td>
<td>7.073</td>
<td>7.071</td>
<td>7.032</td>
<td>7.018</td>
<td>7.013</td>
<td>7.017</td>
<td>7.012*</td>
</tr>
<tr>
<td>SIC</td>
<td>7.092</td>
<td>7.093</td>
<td>7.099</td>
<td>7.068</td>
<td>7.062*</td>
<td>7.065</td>
<td>7.077</td>
<td>7.080</td>
</tr>
</tbody>
</table>

These results are typical of many applications. AIC has not yet reached a minimum value by 8 lags, HIC suggests the choice of 8 lags is optimal, whereas the minimum value of the SIC statistic is $SIC = 7.062$, corresponding to a lag structure with $p = 5$. The AIC criterion is well known to favour longer lag lengths, whereas SIC imposes a higher penalty on additional lags and therefore usually chooses shorter optimal lag lengths. The present results provide empirical evidence that suggest the choice of a lag length $p = 6$ in this example is reasonable.
4.6 Analysing VARs

4.6.1 Granger Causality Testing

In a VAR model, all lagged variables are assumed to contribute information in determining the behaviour of each dependent variable. But in most empirical applications of VARs there are often large numbers of estimated coefficients which are statistically insignificant. A question of considerable importance in empirical work is whether the coefficients of all the lagged values of a particular explanatory variable in a given equation are zero or not. This question bears on whether the information content of the past values of one variable influences the behaviour of another variable in the system. This notion has a close connection with that of causal influence in the sense that predictions might be improved by measuring and including such influences.

In the VAR given in (4.18), for example, the information content of variable $y_2$ on variable $y_1$ might be tested by considering the joint restrictions

$$\phi_{12,1} = \phi_{12,2} = \phi_{12,3} = \cdots = \phi_{12,p} = 0.$$ 

These restrictions on the coefficients of the lagged variables $y_{2t-1}, \ldots, y_{2t-p}$ can be tested jointly using a $\chi^2$ test with $p$ degrees of freedom.

If $y_{2t}$ plays a role in predicting future values of $y_{1t}$, then $y_{2t}$ is said to cause $y_{1t}$ in Granger’s sense (Granger, 1969). It is important to remember that Granger causality is based on the presence (or absence) of predictability and does not of itself signify causal influence. On the other hand, if a causal influence from a certain variable is present, then it is to be expected that such a variable will play a role in prediction. Evidence of Granger causality and the lack of Granger causality from $y_{2t}$ to $y_{1t}$, are denoted, respectively, as

$$y_{2t} \rightarrow y_{1t} \quad y_{2t} \not\rightarrow y_{1t}.$$ 

It is also possible to test for Granger causality in the reverse direction by performing a joint test of the lags of $y_{1t}$ in the $y_{2t}$ equation. Combining both sets of causality results can yield a range of statistical causal patterns:

- **Unidirectional:**
  - **from $y_{1t}$ to $y_{2t}$**
    - $y_{1t} \rightarrow y_{2t}$
    - $y_{2t} \not\rightarrow y_{1t}$

- **Unidirectional:**
  - **from $y_{2t}$ to $y_{1t}$**
    - $y_{2t} \rightarrow y_{1t}$
    - $y_{1t} \not\rightarrow y_{2t}$

- **Bidirectional:**
  - **feedback**
    - $y_{2t} \rightarrow y_{1t}$
    - $y_{1t} \rightarrow y_{2t}$

- **Independence:**
  - $y_{2t} \not\rightarrow y_{1t}$
  - $y_{1t} \not\rightarrow y_{2t}$
Table 4.2 gives results of the Granger causality tests based on the $\chi^2$ statistic in a fitted bivariate VAR(6) model for equity and dividend returns, $re_t$ and $rd_t$, respectively. Both $p$ values are less than 0.05 showing that there is bidirectional Granger causality between equity and dividend returns ($re_t$ and $rd_t$). The results of the Granger causality tests reported in Table 4.2 may be corroborated using estimates from the univariate model in which equity returns are formulated to depend on lags 1 to 6 of both $re_t$ and $rd_t$. In this formulation, a test of the information value of dividend returns in explaining equity returns is given by the statistic $\chi^2 = 20.075$. Since there are 6 degrees of freedom, the $p$ value is 0.003, which confirms that dividend returns are statistically relevant in explaining equity returns at the 5% level. This is consistent with many models in finance which emphasise the information content of dividends in determining equity returns.

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Chi-square</th>
<th>Degrees of Freedom</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$rd \rightarrow re$</td>
<td>20.075</td>
<td>6</td>
<td>0.003</td>
</tr>
<tr>
<td>$re \rightarrow rd$</td>
<td>68.279</td>
<td>6</td>
<td>0.000</td>
</tr>
</tbody>
</table>

4.6.2 Impulse Response Analysis

Granger causality testing is one method of simplifying the system dynamics of a VAR that enhances understanding of variable interactions over time. An alternative, but related approach, focuses on impulse responses by tracking the transmission effects of shocks to the system on the dependent variables. In studying these effects, the dynamic interactions of the system are revealed, showing how the variables respond to shocks over time in relation to one another. This approach to examining system dynamics is called impulse response analysis.

The natural candidate for the shocks in a VAR system is the vector of disturbances $u_t = \{u_{t1}, u_{t2}, ..., u_{tk}\}$, which represents contributions to the dependent variable that are not predicted from past information. The primary problem in the direct use of the disturbance terms in studying impulse responses is that these terms are correlated because $\Omega$ in equation (4.22) contains non-zero off-diagonal elements, which complicates the interpretation of the shocks $u_t$ in terms of underlying economic and financial forces. The solution is to transform the VAR into a new system in which the disturbances are indepen-
dent of each other, commonly referred to as being orthogonal. For a bivariate model the VAR(1) is transformed as

\[
\begin{align*}
y_{1t} &= \alpha_1 + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + v_{1t} \\
y_{2t} &= \alpha_2 + \beta_{20}y_{1t} + \beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + v_{2t}.
\end{align*}
\]

In this system the disturbances satisfy \(E(v_{1t}v_{2t}) = 0\). In specifying this model, a particular ordering has been imposed in which \(y_{1t}\) affects \(y_{2t}\), but \(y_{2t}\) only affects \(y_{1t}\) with a lag. For details of this construction and many other aspects of the identifying procedures involved in the calculation of impulse responses, see Lütkepohl (2005).

Figure 4.4: Impulse responses for the VAR(6) model of equity and dividend returns. The shaded areas represent 95% confidence intervals. Data are monthly for the period January 1871 to September 2016.

Figure 4.5 shows the effects of shocks to bivariate VAR(6) equity-dividend model. These impulse response functions were generated based the assumption that \(re_t\) affects \(rd_t\) contemporaneously, but \(rd_t\) affects \(re_t\) with a lag. In
other words, equity return shocks instantaneously affect both equity returns and dividend returns, whereas dividend return shocks do not have an immediate impact on $r_{et}$.

Four graphics are shown in Figure 4.5 to capture four different sets of impulses. The first column of the figure shows responses to a shock in equity returns and the second column shows responses to a shock in dividends. A positive shock to equity returns has a damped oscillatory effect on itself which quickly dissipates. The effect on dividend returns is initially negative but quickly turns positive and reaches a peak after 8 months, before decaying monotonically. The effect of a positive shock to dividend returns on dividend returns slowly dissipates, approaching zero after nearly 30 periods. The immediate effect of this shock on equity returns is zero by construction in the first period, due to the ordering assumption in this exercise. Subsequently, the effects exhibit a damped oscillatory pattern before decaying rapidly.

### 4.6.3 Variance Decomposition

Impulse response graphics of the type just displayed show the time-forms of system variable responses to incoming shocks. These graphics impart interpretable information about the internal dynamics within a VAR system that govern the transmission effects of shocks. To gain explicit quantitative insight on the relative importance of various shocks on the variables in the system a variance decomposition can also be performed. In this additional analysis, the forecast variances for each variable over the horizon are decomposed into the separate relative effects of each primitive shock with the results expressed as a percentage of the overall movement. The constructive process in this decomposition relies on the same general principles used in analysing impulse responses, so that a recursive structure (or other identifying information) is employed to identify orthogonal shocks associated with each variable in the system. For additional details on the computation of the variance decomposition, see Martin, Hurn and Harris (2013, pp 498–500).

To illustrate, consider the bivariate VAR(6) model for $r_{et}$ and $rd_{t}$, estimated using monthly United States data for the period February 1871 to September 2016, whose parameter estimates are reported in Table 4.1. The 30-period variance decomposition of the VAR, based on the same contemporaneous ordering of variables as the impulse responses, is reported in Table 4.3.

Evidently, dividend shocks contribute very little to equity returns with the maximum contribution being less than 2%. In contrast, equity return shocks after 15 periods contribute more than 10% of the variance in dividend returns. These results suggest that the effects of shocks to equity on dividends are relatively more important than in the reverse direction.
4.7 Modelling Interest Rate Movements

In Chapter 1 the yield curve on government bonds was defined to be a plot of the term structure of yield to maturity against time to maturity at a specific time. An upward sloping yield curve means that the spread between longer-term bond yield and a short-term bond yield is positive. This upward slope is often taken as evidence that short-term bond yields are likely to rise in the future. This conjecture can be assessed by using a test of the hypothesis $\beta > 0$ in the simple regression model

$$r_{1t} - r_{1t-1} = \alpha + \beta (r_{nt} - r_{1t-1}) + u_t,$$  \hspace{1cm} (4.26)

where $r_{1t}$ is the short-term yield on a 1-month bond, $r_{nt}$ is the long-term yield on bonds with maturities greater than one month, and $u_t$ is a disturbance term.

The dynamic regression equation in (4.26) is estimated using yields on United States zero coupon bonds from July 2001 to September 2010. The bond maturities are $\{1, 3, 6, 12, 24, 36, 60, 84, 120\}$ months, with the short-term rate chosen as the 1-month yield. The results are presented in Table 4.4. The first column gives the long-term yield ($n$) used to define the spread relative to the short-term 1-month rate. Inspection of the point estimates shows that the spread is statistically significant in predicting future movements in the 1-month yield in most cases. The strongest statistical relationship is where the spread is computed using the $n = 3$ month bond. The strength of this relationship does diminish as the maturity spectrum widens, remaining statistically significant for the 3-year spread and becoming insignificant thereafter.

An alternate way to conduct a test of the predictive power of the spread for changes in short-term yields is to specify a bivariate VAR containing the change in the 1-month yield $y_{1t} = r_{1t} - r_{1t-1}$ and the spread $y_{2t} = r_{nt} - r_{1t}$, with the

<table>
<thead>
<tr>
<th>Period</th>
<th>Decomposition of $re$</th>
<th>Decomposition of $rd$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$re$</td>
<td>$rd$</td>
</tr>
<tr>
<td>1</td>
<td>100.000</td>
<td>0.000</td>
</tr>
<tr>
<td>5</td>
<td>99.057</td>
<td>0.943</td>
</tr>
<tr>
<td>10</td>
<td>98.785</td>
<td>1.215</td>
</tr>
<tr>
<td>15</td>
<td>98.732</td>
<td>1.267</td>
</tr>
<tr>
<td>20</td>
<td>98.694</td>
<td>1.306</td>
</tr>
<tr>
<td>25</td>
<td>98.680</td>
<td>1.320</td>
</tr>
<tr>
<td>30</td>
<td>98.675</td>
<td>1.325</td>
</tr>
</tbody>
</table>
Table 4.4
Tests of the relationship between yields and spreads based on equation (4.26), using United States bond yields from July 2001 to September 2010.

<table>
<thead>
<tr>
<th>Maturity (months, n)</th>
<th>$\alpha$</th>
<th>se</th>
<th>$\beta$</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-0.121</td>
<td>0.025</td>
<td>1.207</td>
<td>0.187</td>
</tr>
<tr>
<td>6</td>
<td>-0.143</td>
<td>0.032</td>
<td>0.499</td>
<td>0.103</td>
</tr>
<tr>
<td>12</td>
<td>-0.167</td>
<td>0.036</td>
<td>0.378</td>
<td>0.079</td>
</tr>
<tr>
<td>24</td>
<td>-0.150</td>
<td>0.038</td>
<td>0.180</td>
<td>0.046</td>
</tr>
<tr>
<td>36</td>
<td>-0.125</td>
<td>0.040</td>
<td>0.100</td>
<td>0.035</td>
</tr>
<tr>
<td>60</td>
<td>-0.091</td>
<td>0.044</td>
<td>0.041</td>
<td>0.026</td>
</tr>
<tr>
<td>84</td>
<td>-0.077</td>
<td>0.045</td>
<td>0.025</td>
<td>0.021</td>
</tr>
<tr>
<td>120</td>
<td>-0.068</td>
<td>0.047</td>
<td>0.017</td>
<td>0.019</td>
</tr>
</tbody>
</table>

test of predictive ability based on performing Granger causality tests. This strategy has much in common with the dynamic regression equation specified in (4.26) as causality testing involves determining whether lagged values of the spread Granger cause the change in the short-term rate. However, there is one fundamental difference between the two approaches and that is Granger causality tests condition on own lags by including as regressors lags of the dependent variable. To make the two approaches comparable would simply involve augmenting equation (4.26) to include the lags of the dependent variable.

The results of performing Granger causality tests using a bivariate VAR are given in Table 4.5 for alternative spread maturities. The lag length of the VAR is determined optimally using information criteria, with the optimal length reported in the last column. The results of testing for Granger causality in the reverse direction are also reported. The results in the column headed $y_2 \rightarrow y_1$ of Table 4.5 provide strong evidence the spread is a significance predictor of future changes in the 1-month rate as there is Granger causality running from spreads ($y_2$) to the change in the short rate ($y_1$). This result occurs for all maturities with the exception of 10 year bonds where the causality testing produces a $p$ value of 0.124. Interestingly, the empirical results also provide evidence of bidirectional causality running in the opposite direction from the change in the short rate to the spread. This link is statistically significant at 5% level for spreads based on 3-month and 6-month yields and for the longest term spread based on 10-years. There are weaker links at the 10% level for spreads based on 1-year, 3-year, 5-year and 7-year maturities, but not for the spread based on the 2-year yield where there is no evidence of Granger causality.

While the results of the Granger causality tests provide strong evidence that spreads are able to predict future movements in short-term yields, they do
4.7. MODELLING INTEREST RATE MOVEMENTS

Table 4.5

Granger causality tests of the relationship between yields and spreads based on a bivariate VAR containing $y_{1t} = r_{1l} - r_{1l-1}$ and $y_{2t} = r_{nt} - r_{1l}$, using United States bond yields from July 2001 to September 2010. The chi-square statistic is reported together with the corresponding $p$ value.

<table>
<thead>
<tr>
<th>Maturity (months, n)</th>
<th>$y_2 \rightarrow y_1$</th>
<th>$y_1 \rightarrow y_2$</th>
<th>Optimal Lag of the VAR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2$</td>
<td>$p$</td>
<td>$\chi^2$</td>
</tr>
<tr>
<td>3</td>
<td>69.896</td>
<td>0.000</td>
<td>25.080</td>
</tr>
<tr>
<td>6</td>
<td>60.889</td>
<td>0.000</td>
<td>14.090</td>
</tr>
<tr>
<td>12</td>
<td>45.499</td>
<td>0.000</td>
<td>6.375</td>
</tr>
<tr>
<td>24</td>
<td>35.143</td>
<td>0.000</td>
<td>5.931</td>
</tr>
<tr>
<td>36</td>
<td>25.960</td>
<td>0.000</td>
<td>6.806</td>
</tr>
<tr>
<td>60</td>
<td>13.485</td>
<td>0.004</td>
<td>7.300</td>
</tr>
<tr>
<td>84</td>
<td>9.072</td>
<td>0.028</td>
<td>7.394</td>
</tr>
<tr>
<td>120</td>
<td>5.753</td>
<td>0.124</td>
<td>8.210</td>
</tr>
</tbody>
</table>

not provide evidence on the direction of these movements. Because the specification of VARs of the term structure involves estimating a number of parameters, inspection of the parameter estimates does not necessarily help in identifying the direction of these changes. One solution is to compute the impulse response function by shocking the system in terms of a change in the spread and determining the sign of future movements in the short term yield from the estimated impulse responses.

The results of the impulse response analysis are given in Figure 4.5, which plots cumulated impulse responses to a shock in the spread in order to present the results in terms of the actual level of the 1-month yield. The time period of the estimated responses is chosen as 1 year. As there are 4 impulse response plots for each bivariate VAR only those impulse responses showing the effects on the 1-month yield from a positive shock in the spread for each VAR are presented.

The impulse responses show that a positive shock which widens the spreads has a significant positive effect on the 1-month rate. These results are consistent with the view that a positive spread (long-term rates are higher than short-term rates) lead to an increase in the short-term rate. A variation of the regression equation in (4.26) is developed in the exercises. An even more general framework that also nests the bivariate VAR testing strategy is investigated in Chapter 6 where tests of the term structure model are based on the long-run time series properties of interest rates known as cointegration.
Figure 4.5: Impulse responses for the bivariate VAR model of changes in yields and spreads. The plots present the response of the 1-month yield (rather than the change in the 1-month yield) to a shock in the spread. The shaded areas represent 95% confidence intervals. Data are monthly for the period July 2001 to September 2010.

4.8 Diebold-Yilmaz Spillover Index

An important application of the variance decomposition in a VAR model is the spillover index proposed by Diebold and Yilmaz (2009). The objective in constructing this index is to calculate the total contribution of shocks on an asset market arising from the other variables in the VAR. Table 4.6 gives the volatility decomposition for a 10 week horizon of the weekly asset log returns of 19 countries based on a VAR with 2 lags and a constant. The sample period is from 4 December 1996 to 23 November 2007.

The first row of the table gives the contributions to the 10-week forecast variance of shocks in all 19 asset markets on United States weekly returns. By excluding own market shocks, which equal 93.6%, the total contribution of the other 18 asset markets is given in the first cell of the last column headed (Oth-
Table 4.6

Diebold-Yilmaz spillover index of global stock market log returns. Based on a variance decomposition (expressed as a percentage) from estimating a 19 variable VAR with 2 lags and a constant.

<table>
<thead>
<tr>
<th>To</th>
<th>US</th>
<th>UK</th>
<th>FRA</th>
<th>GER</th>
<th>HKG</th>
<th>JPN</th>
<th>AUS</th>
<th>IDN</th>
<th>KOR</th>
<th>MYS</th>
<th>PHL</th>
<th>TAI</th>
<th>THA</th>
<th>ARG</th>
<th>BRA</th>
<th>CHL</th>
<th>MEX</th>
<th>TUR</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>US</td>
<td>93.6</td>
<td>1.6</td>
<td>1.5</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.5</td>
<td>0.3</td>
<td>64</td>
</tr>
<tr>
<td>UK</td>
<td>40.3</td>
<td>55.7</td>
<td>0.7</td>
<td>0.4</td>
<td>0.1</td>
<td>0.5</td>
<td>0.1</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td>FRA</td>
<td>38.3</td>
<td>21.7</td>
<td>37.2</td>
<td>0.1</td>
<td>0.0</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
</tr>
<tr>
<td>GER</td>
<td>40.8</td>
<td>15.9</td>
<td>13.0</td>
<td>27.6</td>
<td>0.1</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.6</td>
<td>0.1</td>
<td>0.3</td>
<td>0.3</td>
<td>0.0</td>
<td>0.2</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>HKG</td>
<td>15.3</td>
<td>8.7</td>
<td>1.7</td>
<td>1.4</td>
<td>69.9</td>
<td>0.3</td>
<td>0.0</td>
<td>0.1</td>
<td>0.0</td>
<td>0.3</td>
<td>0.1</td>
<td>0.0</td>
<td>0.2</td>
<td>0.9</td>
<td>0.3</td>
<td>0.0</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>JPN</td>
<td>12.1</td>
<td>3.1</td>
<td>1.8</td>
<td>0.9</td>
<td>2.3</td>
<td>77.7</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>AUS</td>
<td>23.2</td>
<td>6.0</td>
<td>1.2</td>
<td>0.6</td>
<td>2.3</td>
<td>56.8</td>
<td>0.1</td>
<td>0.4</td>
<td>0.2</td>
<td>0.0</td>
<td>0.2</td>
<td>0.4</td>
<td>0.5</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
<td>0.6</td>
<td>0.7</td>
<td>43.2</td>
</tr>
<tr>
<td>IDN</td>
<td>6.0</td>
<td>1.6</td>
<td>1.2</td>
<td>0.7</td>
<td>6.4</td>
<td>1.6</td>
<td>0.4</td>
<td>77.0</td>
<td>0.7</td>
<td>0.4</td>
<td>0.1</td>
<td>0.9</td>
<td>0.2</td>
<td>1.0</td>
<td>0.7</td>
<td>0.1</td>
<td>0.3</td>
<td>0.1</td>
<td>23.0</td>
</tr>
<tr>
<td>KOR</td>
<td>8.3</td>
<td>2.6</td>
<td>1.3</td>
<td>0.7</td>
<td>5.6</td>
<td>3.7</td>
<td>1.0</td>
<td>1.2</td>
<td>72.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
<td>1.3</td>
<td>0.2</td>
<td>0.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.7</td>
</tr>
<tr>
<td>MYS</td>
<td>4.1</td>
<td>2.2</td>
<td>0.6</td>
<td>1.3</td>
<td>10.5</td>
<td>1.5</td>
<td>0.4</td>
<td>6.6</td>
<td>0.5</td>
<td>69.2</td>
<td>0.1</td>
<td>0.1</td>
<td>0.2</td>
<td>1.1</td>
<td>0.1</td>
<td>0.6</td>
<td>0.4</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>PHL</td>
<td>11.1</td>
<td>1.6</td>
<td>0.3</td>
<td>0.2</td>
<td>8.1</td>
<td>0.4</td>
<td>0.9</td>
<td>7.2</td>
<td>0.1</td>
<td>2.9</td>
<td>62.9</td>
<td>0.3</td>
<td>0.4</td>
<td>1.5</td>
<td>1.6</td>
<td>0.1</td>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>SGP</td>
<td>16.8</td>
<td>4.8</td>
<td>0.6</td>
<td>0.9</td>
<td>18.5</td>
<td>1.3</td>
<td>0.4</td>
<td>3.2</td>
<td>1.6</td>
<td>3.6</td>
<td>1.7</td>
<td>43.1</td>
<td>0.3</td>
<td>1.1</td>
<td>0.8</td>
<td>0.5</td>
<td>0.1</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>TAI</td>
<td>6.4</td>
<td>1.3</td>
<td>1.2</td>
<td>1.8</td>
<td>5.3</td>
<td>2.8</td>
<td>0.4</td>
<td>0.4</td>
<td>2.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9</td>
<td>73.6</td>
<td>0.4</td>
<td>0.8</td>
<td>0.3</td>
<td>0.1</td>
<td>0.3</td>
<td>26.4</td>
</tr>
<tr>
<td>THA</td>
<td>6.3</td>
<td>2.4</td>
<td>1.0</td>
<td>0.7</td>
<td>7.8</td>
<td>0.2</td>
<td>0.8</td>
<td>7.6</td>
<td>4.6</td>
<td>4.0</td>
<td>2.3</td>
<td>2.2</td>
<td>0.3</td>
<td>582.0</td>
<td>0.5</td>
<td>0.2</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>ARG</td>
<td>11.9</td>
<td>2.1</td>
<td>0.6</td>
<td>0.1</td>
<td>1.3</td>
<td>0.8</td>
<td>1.3</td>
<td>0.4</td>
<td>0.4</td>
<td>0.6</td>
<td>0.4</td>
<td>0.6</td>
<td>1.1</td>
<td>0.2</td>
<td>75.3</td>
<td>0.1</td>
<td>0.1</td>
<td>1.4</td>
<td>0.3</td>
</tr>
<tr>
<td>BRA</td>
<td>14.1</td>
<td>1.3</td>
<td>1.0</td>
<td>0.7</td>
<td>1.3</td>
<td>1.4</td>
<td>1.6</td>
<td>0.5</td>
<td>0.5</td>
<td>0.7</td>
<td>1.0</td>
<td>0.8</td>
<td>0.1</td>
<td>0.7</td>
<td>7.1</td>
<td>65.8</td>
<td>0.1</td>
<td>0.6</td>
<td>0.7</td>
</tr>
<tr>
<td>CHL</td>
<td>11.8</td>
<td>1.1</td>
<td>1.0</td>
<td>0.0</td>
<td>3.2</td>
<td>0.6</td>
<td>1.4</td>
<td>2.3</td>
<td>0.3</td>
<td>0.3</td>
<td>0.1</td>
<td>0.9</td>
<td>0.3</td>
<td>0.8</td>
<td>2.9</td>
<td>4.0</td>
<td>65.8</td>
<td>2.7</td>
<td>0.4</td>
</tr>
<tr>
<td>MEX</td>
<td>22.2</td>
<td>3.5</td>
<td>1.2</td>
<td>0.4</td>
<td>3.0</td>
<td>0.3</td>
<td>0.3</td>
<td>1.2</td>
<td>0.2</td>
<td>0.3</td>
<td>0.9</td>
<td>1.0</td>
<td>0.1</td>
<td>0.3</td>
<td>0.5</td>
<td>5.4</td>
<td>16.0</td>
<td>0.3</td>
<td>56.9</td>
</tr>
<tr>
<td>TUR</td>
<td>3.0</td>
<td>2.5</td>
<td>0.2</td>
<td>0.7</td>
<td>0.6</td>
<td>0.9</td>
<td>0.6</td>
<td>0.1</td>
<td>0.6</td>
<td>0.3</td>
<td>0.6</td>
<td>0.1</td>
<td>0.9</td>
<td>0.8</td>
<td>0.5</td>
<td>1.1</td>
<td>0.6</td>
<td>0.2</td>
<td>85.8</td>
</tr>
</tbody>
</table>

Others 291.9 84.1 31 11.2 80.8 19.2 11.5 31.4 13.6 16.2 9.9 8.2 5.9 11.8 21.4 9.4 2.6 8.4 6.7 67.1 675.1

Own 385.5 139.8 68.2 38.8 150.6 96.9 68.3 108.3 86.4 85.4 72.8 51.2 79.5 70.0 96.7 75.2 68.4 65.4 92.4

Index = 35.5%
1.6 + 1.5 + · · · + 0.3 = 6.4%.

Similarly, for the United Kingdom in the second row of the table, the total contribution of the other 18 asset markets to its forecast variance is

40.3 + 0.7 + · · · + 0.5 = 44.3%.

Of the 19 asset markets, the United States appears to be the most independent of all international asset markets as it has the lowest contributions from all other markets, equal to just 6.4%. The next lowest is Turkey with a contribution of 14.2%. Germany’s asset market appears to be the most affected by international shocks with shocks from external markets contributing 72.4% to its forecast variance.

Adding the separate contributions to each asset market in the last column gives the total contributions of external (to own market) shocks on all 19 asset markets

6.4 + 44.3 + · · · + 14.2 = 675.1%.

As the contributions to the total forecast variance by construction are normalised to sum to 100% for each of the 19 asset markets, the percentage contribution of external shocks to these 19 asset markets is given by the spillover index

\[
SPILLOVER = \frac{675.1}{19} = 35.5%.
\]

This value shows that on average approximately one-third of the forecast variance of asset returns in these countries is the result of foreign shocks with the remaining two-thirds arising from domestic shocks.

4.9 Exercises

The data required for the exercises are available for download as EViews workfiles (*.wf1), Stata datafiles (*.dta), comma delimited text files (*.csv) and as Excel spreadsheets (*.xlsx).

1. Equity and Dividend Returns

The data are monthly observations on United States equity prices and dividends for the period January 1871 to September 2016.
(a) Compute the percentage monthly log returns on equities and dividends expressed as percentages, defined as

\[ re_t = 100 \times (\log P_t - \log P_{t-1}) \]
\[ rd_t = 100 \times (\log D_t - \log D_{t-1}). \]

Plot the two series and interpret their time series patterns.

(b) Compute the ACF of equity returns for up to 6 lags. Compare a manual procedure with an automated version provided by econometric software.

(c) Compute the PACF of equity returns for up to 6 lags. Compare a manual procedure with an automated version provided by econometric software.

(d) Repeat parts (b) and (c) for dividend returns.

(e) Estimate an AR(6) model of equity returns. Interpret the parameter estimates.

(f) Estimate an MA(3) model of equity returns, using the following two-step method (Durbin, 1959) to circumvent the necessity of using a nonlinear estimation procedure. The steps are as follows.

(i) Estimate an autoregressive model for \( y_t \) with a constant term and \( P > 3 \) autoregressive terms,

\[ re_t = \phi_0 + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-P} + u_t \]

and save the residuals \( \hat{u}_t \).

(ii) Estimate the MA(3) parameters by estimating the model

\[ re_t = \psi_0 + \psi_1 \hat{u}_{t-1} + \psi_2 \hat{u}_{t-2} + \psi_3 \hat{u}_{t-3} + v_t, \]

by ordinary least squares.

Comment on the results.

(g) Using the same two-step procedure, estimate an ARMA(1,1) model of equity returns. The procedure is now as follows.

(i) Estimate an AR(P) model with \( P \) and save the residuals, \( \hat{u}_t \).

(ii) Estimate an ARMA(\( p, q \)) model using 1 lag of the dependent variable and 1 lags of the fitted residuals, \( \hat{u}_{t-1} \). Interpret your results.

2. Forward Market Efficiency
The data are weekly observations (all recorded on a Wednesday) for the period 4 January 1984 to 31 December 1990 on the spot U.S./Australian exchange rate and the 1-month, 3-month and 6-month forward U.S./Australian exchange rates.

(a) The forward market is efficient if the lagged forward rate is an unbiased predictor of the current spot rate. Estimate the following model of the spot and the lagged 1-month forward rate

\[ s_t = \beta_0 + \beta_1 f_{t-4} + u_t, \]

where the forward rate is lagged four periods (the data are weekly). Test the restriction \( \beta_1 = 1 \) and interpret the result.

(b) Compute the ACF and PACF of the least squares residuals, \( \hat{u}_t \), for the first 8 lags. Interpret the results.

(c) Repeat parts (a) and (b) for the 3-month and the 6-month forward rates.

3. Mean Aversion and Reversion in Stock Returns

The data files contain, respectively, annual, monthly and weekly log returns data (expressed as percentages) for the years 1989 to 2009 on the Australian share index, the NASDAQ index and the Singapore Straits Times index.

(a) Estimate the following regression equation using returns on the NASDAQ, \( r_{et} \), for each frequency (monthly, quarterly, annual)

\[ r_{et} = \phi_0 + \phi_1 r_{et-1} + u_t, \]

where \( u_t \) is a disturbance term. Interpret the results.

(b) Repeat part (a) for the Australian share price index.

(c) Repeat part (a) for the Singapore Straits Times stock index.

4. An Equity-Dividend VAR
4.9. EXERCISES

Use the same data set as in Exercise 1 and also the same definitions of equity, \( r e_t \), and dividend, \( r d_t \), returns.

(a) Compute the percentage monthly returns on equities and dividends and estimate a bivariate VAR(6) for these variables.

(b) Test for the optimum choice of lag length using information criteria and specifying a maximum lag length of 12.

(c) Test for Granger causality between equities and dividends and interpret the results.

(d) Compute the impulse responses for 30 periods and interpret the results.

(e) Compute the variance decomposition for 30 periods and interpret the results.

5. Campbell-Shiller Present Value Model

The data are monthly data for the period January 1933 to December 1990 comprising observations on United States equity prices and dividend payments. Let \( y_t \) be dividend yields (expressed in percentage terms) and let \( v_t \) be deviations from the present value model

\[
p_t = \beta + \alpha d_t + v_t,
\]

where \( p_t \) is the log of equity prices and \( d_t \) log dividend payments. Campbell and Shiller (1987) develop a VAR(1) model for \( y_t \) and \( v_t \) given by

\[
\begin{bmatrix}
y_t \\
v_t
\end{bmatrix} = \begin{bmatrix}
\varphi_{10} & \varphi_{11} \\
\varphi_{20} & \varphi_{21}
\end{bmatrix} \begin{bmatrix}
y_{t-1} \\
v_{t-1}
\end{bmatrix} + \begin{bmatrix}
u_{1t} \\
u_{2t}
\end{bmatrix}.
\]

(a) Estimate the parameter \( \alpha \) and compute the least squares residuals \( \hat{v}_t \).

(b) Estimate the VAR(1) containing the dividend yields, \( y_t \), and \( \hat{v}_t \).

(c) Campbell and Shiller show that

\[
\varphi_{22} = \delta^{-1} - \alpha \varphi_{12},
\]

where \( \delta \) represents the discount factor. Use the parameter estimate of \( \alpha \) obtained in part (a) and the parameter estimates of \( \varphi_{12} \) and \( \varphi_{22} \) obtained in part (b), to estimate \( \delta \). Interpret the result.

6. Campbell-Shiller Model of Interest Rates
CHAPTER 4. DYNAMIC MODELLING WITH STATIONARY VARIABLES

The data are monthly observations on United States zero coupon bonds with maturities of \{1, 3, 6, 12, 24, 36, 60, 84, 120\} months. The sample period is from July 2001 to September 2010. Campbell and Shiller (1991) proposed testing the relationship between short-term rates of maturity \(m\) and longer-term rates of maturity \(n > m\) using the regression.

\[
r_{n-m,t} - r_{n,t-m} = \alpha + \beta(r_{n,t-m} - r_{m,t-m}) + u_t,
\]

where \(u_t\) is a disturbance term. Of particular interest is whether or not the slope coefficient \(\beta\) satisfies the relationship \(\beta = m / (n - m)\).

(a) Test this relationship for maturities of \(m = 3\) months and \(n = 6\) months.
(b) Test this relationship for maturities of \(m = 1\) year and \(n = 2\) years.
(c) Test this relationship for maturities of \(m = 1\) year and \(n = 3\) years.

7. Diebold-Yilmaz Spillover Index

Diebold and Yilmaz (2009) construct spillover indices of international real asset returns and volatility based on the variance decomposition of a VAR. The data file contains weekly data on real asset returns, \(rets\), and volatility, \(vol\), of 7 developed countries and 12 emerging countries from the first week of January 1992 to the fourth week of November 2007.

(a) Compute descriptive statistics of the 19 real asset market returns given in \(rets\). Compare the estimates with the results reported in Table 1 of Diebold and Yilmaz.
(b) Estimate a VAR(2) containing a constant and the 19 real asset market returns.
(c) Estimate \(VD_{10}\), the variance decomposition for horizon \(h = 10\), and compare the estimates with the results reported in Table 3 of Diebold and Yilmaz.
(d) Using the results in part (c) compute the ‘Contribution from Others’ by summing each row of \(VD_{10}\) excluding the diagonal elements, and the ‘Contribution to Others’ by summing each column of \(VD_{10}\) excluding the diagonal elements. Interpret the results.
(e) Repeat parts (a) to (d) with the 19 series in \(rets\) replaced by \(vol\), and the comparisons now based on Tables 2 and 4 in Diebold and Yilmaz.
Chapter 5

Nonstationarity in Financial Time Series

An important property of asset prices identified in Chapter 2 is that they exhibit strong evidence of trends over long periods of time. Trend behaviour often manifests in a tendency for a time series to drift over time in such a way that no fixed mean value is revealed. This property is highlighted in Figure 2.1 which shows that the S&P 500 stock price index displays a general pattern of exponential growth over a long historical period from 1950 to 2016. But this long term growth is coupled with extended sub-periods in which prices wander above and below the growth line. Such time series are said to be nonstationary and may embody both a deterministic drift, which manifests in the exponential growth path shown in Figure 2.1, and a stochastic process trend, which arises from the accumulation of random forces that drive prices to wander above and below the path of deterministic drift.

Nonstationarity in this complex form that combines deterministic and random forces is one of the primary characteristics of financial time series. Nonstationary behaviour needs to be respected in empirical work because of the importance of linkages between trending financial time series that are often the subject of investigation, because of the serious impact that trends can have on forecasting performance, and because of major changes in the econometric apparatus of inference when trends are present in the data.

Financial series that exhibit no such trending behaviour and whose mean, variance and autocovariances are time invariant are said to be stationary and are the subject matter of Chapter 4. The present chapter provides an introduction to some of the main properties of nonstationary time series and the models that generate such series. Its primary focus is on identifying and testing for nonstationarity in financial time series.

Identification of stochastic, as distinct from deterministic, nonstationarity,
typically hinges on testing evidence in support of a unit root restriction $\rho = 1$ in an autoregressive model of the form

$$y_t = \rho y_{t-1} + v_t,$$  \hspace{1cm} (5.1)

in which $v_t$ is a stationary disturbance term. If the restriction $\rho = 1$ is satisfied and $v_t \sim iid (0, \sigma_v^2)$, where from Chapter 4 $iid$ stands for independently and identically distributed, this model is commonly known as a random walk without drift. Tests of the restriction $\rho = 1$ are referred to as unit root tests and have very different characteristics from traditional regression tests in stationary time series models where $|\rho| < 1$. The trajectories of unit root processes may be regarded as random draws of a function that is observed at discrete points in time over the interval of observation. In consequence, under a unit root null hypothesis, estimates of the parameter $\rho$ and test statistics of the hypothesis that $\rho = 1$ have asymptotic distributions that rely on the distribution of these entire trajectories. These distributions differ considerably from a normal distribution and lead to new procedures for testing.

The classification of variables as either stationary or nonstationary has implications in both finance and econometrics. From a finance perspective, stochastic nonstationarity is important because the ubiquity of the random forces driving financial asset prices leads to the wandering price trajectories that are typically observed in practice. Within this class of models, unit root processes like (5.1) with $\rho = 1$ have special significance because they are compatible with common formulations of the efficient markets hypothesis. According to this hypothesis in its general form, all the information about the future price of an asset is embodied in the most recently observed price. So the conditional expectation of tomorrow’s price, given the price history to today, is simply today’s price.

Small departures from the unit root class both above and below unity are also important. If nonstationarity takes the form of an explosive process with $\rho > 1$, then this may be taken as evidence of the emergence of financial exuberance associated with the expansionary phase of a bubble in the price of the asset. For values of $\rho$ less than unity, as already noted, $y_t$ in (5.1) is a stationary process with statistical properties that are very different from the unit root ($\rho = 1$) and explosive ($\rho > 1$) processes.

### 5.1 The Random Walk with Drift

**Specification**

The return to a risky asset in an efficient market may be written as

$$r_t = p_t - p_{t-1} = \alpha + v_t, \hspace{1cm} v_t \sim iid (0, \sigma_v^2),$$  \hspace{1cm} (5.2)
5.1. THE RANDOM WALK WITH DRIFT

where \( p_t \) is the logarithm of the asset price. The parameter \( \alpha \) represents the average return on the asset. From an efficient markets point of view, since \( \alpha = 0 \) and \( v_t \) is not autocorrelated, \( r_{t+1} \) cannot be predicted using information at time \( t \).

Another way of expressing (5.2) is to write the equation in terms of \( p_t \) as

\[
p_t = \alpha + p_{t-1} + v_t. \tag{5.3}
\]

The parameter \( \alpha \) is the drift parameter with \( y_t \) now representing a random walk with drift, so that the dependence of \( p_t \) on the past history of shocks \( v_j \) and the origination of the process is apparent. Start by lagging the random walk with drift model in equation (5.3) by one period, giving

\[
p_{t-1} = \alpha + p_{t-2} + v_{t-1},
\]

and substituting this expression for \( p_{t-1} \) in (5.3) gives

\[
p_t = \alpha + \alpha + p_{t-2} + v_t + v_{t-1}.
\]

Repeating this recursive substitution process for \( t \)-steps gives

\[
p_t = p_0 + \alpha t + v_t + v_{t-1} + v_{t-2} + \cdots + v_1 = \alpha t + \sum_{j=1}^{t} v_j + p_0, \tag{5.4}
\]

in which \( p_0 \) is fully determined by its initial value, a deterministic trend component and the accumulation of the complete history of shocks since initialisation of the process at \( t = 0 \).

The unit root mechanism is evident in the equation both in the unit coefficient of the lagged price variable \( p_{t-1} \) and in the accumulation process \( \sum_{j=1}^{t} v_j \) whose weights are unity in all time periods. The drift parameter \( \alpha \) now determines the extent of the deterministic drift measured by the linear time trend \( \alpha t \). From an efficient market perspective this equation shows that in predicting the price of an asset in the next period, all of the relevant information is contained in the current price when \( \alpha = 0 \).

Properties

Taking expectations of expression (5.4) and using the property that \( E(v_t) = E(v_{t-1}) = \cdots = 0 \), gives the mean of \( p_t \)

\[
E(p_t) = p_0 + \alpha t.
\]

Evidently when \( \alpha > 0 \), the mean price drifts upwards and increases over time at the same constant rate \( \alpha \). Even when the drift parameter \( \alpha \) is small, over long periods of time the upward drift in the mean price becomes a prominent
characteristic of the time series. The variance of $p_t$ is given at each point in time by

$$\text{var}(p_t) = \mathbb{E}\{[p_t - \mathbb{E}(p_t)]^2\} = t\sigma_v^2,$$

which uses the iid property that the component shocks $v_j$ are uncorrelated.

Just as for the mean, the variance is also a linear increasing function over time. So the asset price $p_t$ exhibits greater overall variation, or fluctuations with increasing amplitude, as time passes. These properties reveal some of the implications of the efficient market hypothesis on the time series behaviour of financial asset prices. Specifically, in an efficient market asset prices may be expected to exhibit trending behaviour in levels and in long term fluctuations.

To appreciate the capabilities of a simple time series model such as (5.4) in capturing the main time series features of financial asset prices it is useful to view a simulation of this time series. Figure 5.1 plots a simulated random walk with drift based on equation (5.4). The parameters of the random walk with drift are taken to be $p_0 = 1.491$, $\alpha = 0.0035$ and $\sigma_v^2 = 0.002$.

![Figure 5.1: Simulated trajectory of a random walk with drift based on equation (5.4) with parameter values $p_0 = 1.491$, $\alpha = 0.0035$ and $\sigma_v^2 = 0.002$. The distribution of the disturbance term, $v_t$ in (5.3), is taken to be normal.](image)

Observe that the simulated price exhibits two major characteristics: (i) an increasing mean, associated with the positive drift upwards in the series; and (ii) an increasing variance over time, associated with the accumulation of

---

1These values are based on the S&P 500 index from January 1871 to Sept 2016, where $p_0$ is the log price in January 1871, and $\alpha$ and $\sigma_v^2$ are, respectively, the mean and the variance of log returns over the sample period.
shocks, $\sum_{j=1}^{t} v_{j}$, that appears in the model solution (5.4). The simulated price series has similar characteristics to the observed logarithm of the price index given in Figure 2.2 in Chapter 2.

**Order of Integration**

The partial summation component of $p_{t}$, namely $\sum_{j=1}^{t} v_{j}$ in equation (5.4) is known as a stochastic trend component, which aggregates (or integrates) up the component shocks $v_{j}$. Simply removing the deterministic trend, $\alpha t$, from the process $y_{t}$ will not be sufficient to obtain a stationary series because the stochastic trend component is still retained. This partial summation component is the origin of an important concept concerning nonstationarity, namely, the order of integration of a time series. A process is integrated of order $d$, denoted by $I(d)$, if it can be rendered stationary by differencing $d$ times. Setting $d = 1$ to ensure differencing once, we have

$$\Delta p_{t} = \alpha + v_{t},$$

where the symbol $\Delta$ is known as the difference operator, leading to $\Delta p_{t} = p_{t} - p_{t-1}$. The series $\Delta p_{t}$ is stationary with mean $\alpha$, variance $\sigma^2$, and residual $v_{t}$ which in the case of model (5.3) is serially uncorrelated.

A general time series $y_{t}$ is said to be integrated of order one, denoted $I(1)$, if it is rendered stationary by differencing once: that is $y_{t}$ is nonstationary, but $\Delta y_{t} = y_{t} - y_{t-1}$ is stationary. If $d = 2$, then $y_{t}$ is $I(2)$ and needs to be differenced twice to achieve stationarity as follows

$$\Delta(y_{t} - y_{t-1}) = (y_{t} - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_{t} - 2y_{t-1} + y_{t-2}.$$  

It is rare for financial series to exhibit orders of integration as high as $d = 2$ and extremely rare to encounter even higher orders than $I(2)$. In general, the series may be regarded as integrated of order $d$, or $I(d)$ with $d \geq 1$, if it requires differencing $d$ times to achieve stationarity. By analogy, a stationary process is integrated of order zero, $I(0)$, because it does not require any differencing to achieve stationarity.

5.2 Characteristics of Financial Data

Most financial econometricians agree that the dominant characteristics of many financial time series arise from the random forces involved in stochastic

\[2\] Strictly speaking, the $I(1)$ property also requires that $\Delta y_{t}$ not be representable as a difference of some stationary time series $w_{t}$, in which case $\Delta y_{t} = y_{t} - y_{t-1} = w_{t} - w_{t-1}$ and then $y_{t} = w_{t}$ would itself be stationary.

\[3\] Fractional values of $d$ are also possible and these correspond to more general forms of nonstationarity (sometimes referred to as long-memory) in the data. While these long-memory processes are appropriate in some instances in financial econometric work, attention here is confined to $I(d)$ processes with integer values of $d$. 
process trends rather than deterministic trends. It is particularly hard to reconcile the strong predictability that is implied by a deterministic trend with the complications and surprises that are continually faced by financial forecasters.\footnote{Indeed, it was the astonishing failure of stock market forecasters, including some leading economists such as Irving Fisher, during the period leading up to the great stock market crash of October 1929 that led Alfred Cowles in 1932 to establish the Cowles Commission for Research in Economics whose immediate goal was to advance a scientific understanding of stock market forecasting.}

To illustrate the point about the importance of the stochastic trend component in bond market data, consider fitting a simple AR(1) regression model

\[ y_t = \alpha + \rho y_{t-1} + v_t, \tag{5.5} \]

to bond yields. The empirical results obtained by fitting this regression to monthly data on United States zero coupon bonds with maturities ranging from 2 months to 9 months for the period January 1947 to February 1987 are given in Table 5.1.

<table>
<thead>
<tr>
<th>Maturity (mths)</th>
<th>Intercept $\hat{\alpha}$</th>
<th>$\text{se}(\hat{\alpha})$</th>
<th>Slope $\hat{\rho}$</th>
<th>$\text{se}(\hat{\rho})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.090</td>
<td>0.046</td>
<td>0.983</td>
<td>0.008</td>
</tr>
<tr>
<td>3</td>
<td>0.087</td>
<td>0.045</td>
<td>0.984</td>
<td>0.008</td>
</tr>
<tr>
<td>4</td>
<td>0.085</td>
<td>0.044</td>
<td>0.985</td>
<td>0.007</td>
</tr>
<tr>
<td>5</td>
<td>0.085</td>
<td>0.045</td>
<td>0.985</td>
<td>0.007</td>
</tr>
<tr>
<td>6</td>
<td>0.087</td>
<td>0.045</td>
<td>0.985</td>
<td>0.007</td>
</tr>
<tr>
<td>9</td>
<td>0.088</td>
<td>0.046</td>
<td>0.985</td>
<td>0.007</td>
</tr>
</tbody>
</table>

The primary result of interest in the estimated regressions reported in Table 5.1 is that the fitted slope coefficient, $\hat{\rho}$, is universally close to unity and strongly indicative of the presence of a stochastic trend in the data. A secondary result of importance is that the fitted intercept $\hat{\alpha}$ is small and positive, indicative of a very small (and possibly insignificant) drift in the time series over time, an outcome that is consistent across all maturities and, accordingly, the more persuasive. This general pattern in the estimated coefficients observed in Table 5.1 is indicative of a much more general finding that is almost ubiquitous in other financial markets, such as currency markets (spot and forward exchange rates), equity markets (share prices and dividends), and commodity markets (such as oil, gas, gold, and copper).

As an example of financial data drawn from equity rather than bond markets, consider Figure 5.2 that shows time series of real United States equity prices...
and various transformations of this series over the period from January 1871 to June 2004. Even casual inspection of these plots reveals prominent characteristics of equity prices that relate to the discussion concerning the random walk with drift model and the importance of the stochastic trend component. It is clear from the top row of Figure 5.2 that equity prices and the logarithm of equity prices are nonstationary. Both these series exhibit a general upwards drift in mean, signaling $\alpha > 0$ and confirming a positive trend over this long historical period.

![Equity Prices](image)
![Logarithm of Equity Prices](image)

**Figure 5.2**: Monthly United States equity prices and various transformations of the equity price process for the period January 1871 to June 2004.

Simple first differencing of equity prices, on the other hand, renders the series constant in mean over time but with an apparent nonstationarity in the variance, as the variability of the first differences increases dramatically over time. The implication is that first differencing equity prices does not on its own produce a stationary series. Finally, equity returns defined as the first differences of the logarithms of prices evidently manifest greater stability over time, both in mean and in variance. But considerable fluctuations in the variance are still apparent in the equity returns series, indicating some instability or potential nonstationarity in variance.

The plots shown in Figure 5.2 reveal that certain transformations of time series, such as taking logarithms and first differencing, can be useful in reveal-
The above discussion points to some of their advantages and limitations. In particular, first differencing prices may help to remove nonstationarity in the mean and first differencing logarithms of prices may help to stabilise the variance, which reinforces some of the ideas suggested by considering simple models such as (5.4). While such transformations are helpful in interpreting properties of financial time series and are heavily used in practical work, there are limitations to what they can accomplish. As the plot of equity returns in Figure 5.2 shows, however, there is still much to explain in the apparent time-varying volatility of this series.

5.3 Dickey-Fuller Methods and Unit Root Testing

The methods considered in Dickey and Fuller (1979; 1981) were developed as tests of the hypothesis $\rho = 1$ in equation (5.5) with the additional assumption that the disturbance term, $v_t$, is normally distributed.\(^6\) With some modifications to the form of these tests and under more general assumptions and limit theory, these procedures provide a framework for assessing evidence for the presence of a unit root in quite general time series settings. This framework is now extensively used in applied econometric work and it remains one of the most popular methods of testing for nonstationarity in financial time series. The approach is called unit root testing.

5.3.1 The Dickey-Fuller (DF) Unit Root Tests

Consider again the AR(1) regression equation
\[
y_t = \alpha + \rho y_{t-1} + v_t, \tag{5.6}
\]
in which $v_t \sim N(0, \sigma^2_v)$. The relevant null and alternative hypotheses are stated as follows:

\[
H_0 : \quad \rho = 1 \quad [\text{Variable } y_t \text{ is nonstationary}]
\]
\[
H_1 : \quad |\rho| < 1 \quad [\text{Variable } y_t \text{ is stationary}]. \tag{5.7}
\]

Importantly, $H_1$ is a left-sided alternative to the unit root null $H_0$. To perform the test, equation (5.6) is estimated by ordinary least squares regression and

\(^5\)These transformations are examples of filters. Filters are particularly useful in detrending data and removing recurrent effects such as seasonality. They are therefore heavily used in empirical work.

\(^6\)Extension to more general conditions where the data are not normally distributed and the innovations are serially dependent and possibly heterogeneous involves what is known as an invariance principle because the final asymptotic result is invariant to the assumption of normality. Phillips (1987) verified the existence of such an invariance principle in this setting, showing that the limit results obtained by Dickey and Fuller (1979, 1981) have more general representations and with certain adjustments to the statistics the limit theory remains valid under much more general conditions.
5.3. DICKEY-FULLER METHODS AND UNIT ROOT TESTING

A $t$ statistic is constructed in the usual manner to test whether $\rho = 1$. This statistic has the conventional ratio form

$$t_\rho = \frac{\hat{\rho} - 1}{\text{se}(\hat{\rho})},$$

(5.8)

where $\text{se}(\hat{\rho})$ is the standard error of $\hat{\rho}$. There is no deviation from normal practice up to this point in testing the hypothesis: estimation of (5.6) by ordinary least squares, construction of the standard error of $\hat{\rho}$, and the use of the $t$ statistic in (5.8) to test the hypothesis $H_0$ are all sound procedures. The difficulty in executing the test arises from the fact that under the null hypothesis, the time series $y_t$ is nonstationary and nonstationarity affects both the finite sample and asymptotic distribution of the statistic $t_\rho$. Even under the normality assumption used by Dickey and Fuller (1979, 1981) the statistic $t_\rho$ does not have a $t$ distribution.

In practice, it is convenient to transform equation (5.6) in a way that converts the $t$ statistic in (5.8) to a test of a zero slope coefficient in the transformed equation. This transformation has the great advantage that the $t$ statistic commonly reported in standard regression packages directly yields the unit root test statistic. To achieve this transformation, simply subtract $y_{t-1}$ from both sides of (5.6) and collect terms to give

$$y_t - y_{t-1} = \alpha + (\rho - 1)y_{t-1} + v_t.$$  

(5.9)

Defining $\beta = \rho - 1$ gives the regression equation

$$y_t - y_{t-1} = \alpha + \beta y_{t-1} + v_t.$$  

(5.10)

Equations (5.6) and (5.10) are precisely the same equations with the connection between them being the simple reparametrisation $\beta = \rho - 1$ in (5.10).

To illustrate this equivalence in a practical application, consider the monthly data on United States zero coupon bonds with maturities ranging from 2 months to 9 months for the period January 1947 to February 1987. This data is used in the estimation of the AR(1) regressions reported in Table 5.1. Estimation of equation (5.6) yields the following results (with standard errors given in parentheses)

$$y_t = 0.090 + 0.983 y_{t-1} + \tilde{v}_t,$$  

(5.11)

On the other hand, estimating the transformed equation (5.10) yields

$$y_t - y_{t-1} = 0.090 - 0.017 y_{t-1} + \tilde{v}_t.$$  

(5.12)

Comparing the estimated equations in (5.11) and (5.12) shows that they differ only in terms of the slope estimate on $y_{t-1}$. The difference in the two slope estimates is easily reconciled as the slope estimate of (5.11) is $\hat{\rho} = 0.983$, whereas an estimate of $\hat{\beta}$ may be recovered as

$$\hat{\beta} = \hat{\rho} - 1 = 0.983 - 1 = -0.017.$$
This is also the slope estimate obtained in (5.12). To perform a statistical test of the null hypothesis $H_0 : \rho = 1$, the two relevant $t$ statistics in these two regressions are

\[ t_\rho = \frac{\hat{\rho} - 1}{\text{se}(\hat{\rho})} = \frac{0.983 - 1}{0.008} = -2.120, \]

\[ t_\beta = \frac{\hat{\beta} - 0}{\text{se}(\hat{\beta})} = \frac{-0.017 - 0}{0.008} = -2.120, \]

demonstrating that the two methods are indeed equivalent.

These tests can be extended to deal with the possibility that under the alternative hypothesis the time series may be stationary around a deterministic trend. As discussed in Sections 5.1 and 5.2, financial data typically exhibit trends and empirical researchers therefore face the difficulty of distinguishing between stochastic and deterministic trends. If the data do trend over time and if the null hypothesis of nonstationarity is rejected, it is important that the maintained model under the alternative hypothesis is able to account for the major characteristics displayed by the series being tested. If the regression equation (5.10) is used for testing and the null hypothesis of a unit root is rejected, the alternative hypothesis is that of a process which is stationary around a constant. In other words, the model under the alternative hypothesis contains no deterministic trend. The maintained hypothesis is therefore not general enough to accommodate a trend stationary time series as an alternative. Consequently, an important extension of the unit root testing framework is to include a linear time trend in the test regression to account for any deterministic drift that may be present in the time series.

The form of the equation to be estimated is

\[ y_t - y_{t-1} = \alpha + \beta y_{t-1} + \delta t + v_t. \] (5.13)

The Dickey-Fuller test still consists of testing $\beta = 0$. But under the alternative hypothesis, $y_t$ is now a stationary process with a deterministic trend.

Once again using the monthly data on United States zero coupon bonds, the estimated regression including the time trend gives the following results (with standard errors in parentheses)

\[ \Delta y_t = \begin{align*}
0.030 & -0.046 & y_{t-1} \quad + \quad 0.001 & t \quad + \quad \hat{v}_t.
\end{align*} \]

The value of the Dickey-Fuller test statistic is

\[ t_\beta = \frac{\hat{\beta} - 0}{\text{se}(\hat{\beta})} = \frac{-0.046 - 0}{0.014} = -3.172. \]

Finally, unit root tests can be performed without a constant and a time trend by setting $\alpha = 0$ and $\delta = 0$ in (5.13). This form of the test, which assumes
5.3. DICKEY-FULLER METHODS AND UNIT ROOT TESTING

that the process has zero mean, is mainly used when testing the residuals of a regression for stationarity as these residuals have sample mean zero by construction whenever there is a constant in the regression equation. Residual-based applications of tests for unit roots are considered in Chapter 6.

In summary, three basic forms of the Dickey-Fuller unit root test are available, based on the following simple regression equations

\[
\begin{align*}
\text{Model 1:} & \quad \Delta y_t = \beta y_{t-1} + v_t, \\
\text{Model 2:} & \quad \Delta y_t = \alpha + \beta y_{t-1} + v_t, \\
\text{Model 3:} & \quad \Delta y_t = \alpha + \delta t + \beta y_{t-1} + v_t.
\end{align*}
\]  

For each of these three models, the null hypothesis of the unit root test remains the same, namely, \( H_0 : \beta = 0 \). Unlike conventional statistical testing, however, the pertinent critical value for determining statistical significance in each case is different. The differences arise because the distribution of the unit root test statistic changes substantially depending on which model is used as the test regression. Thus, changing the regression equation by adding an intercept and/or a linear time trend not only affects the fitted regression coefficients and \( t \) statistics, it also changes their asymptotic distributions.

![Graph](image)

Figure 5.3: Comparisons of the simulated centred and scaled distributions of the scaled estimator \( T \hat{\beta} \) for Dickey-Fuller regressions in cases (i) without an intercept or trend (dashed line), (ii) with an intercept but without a trend (long-dashed line) and (iii) with both intercept and trend (solid line).

To illustrate these differences, the distributions of the scaled coefficient statistic \( T \hat{\beta} \) for the different versions of the Dickey-Fuller test regression are shown...
in Figure 5.3. In addition Figure 5.4 shows the distributions of the Dickey-Fuller $t$ statistics for each of the models. The key point to note in these distributions is that all three distributions are heavily skewed to the left and are also located to the left of a standard normal distribution.\footnote{These asymptotic distributions are derived using a body of probability theory known as functional central limit theory. Functional central limit theory is concerned with establishing that certain function-valued random elements have limits as Gaussian stochastic processes, the most famous of which is Brownian motion. Brownian motion was discovered by the British botanist Robert Brown in 1827 after whom the process is named. Its properties were explored and explained for the first time by Albert Einstein in 1905. Somewhat later, the mathematician Norbert Wiener provided its probabilistic foundations in the function space of continuous functions, and for his fundamental contribution Brownian motion is often called the Wiener process.} In addition, these distributions become more dispersed, more located to the left of the origin, and less negatively skewed as more deterministic components (constants and time trends) are included.\footnote{This phenomenon was explained analytically in Phillips (2002).} This skewness has a major impact on statistical inference because the critical values of the test statistic are very different from those of conventional statistical tests with an asymptotic normal distribution. Table 5.2 provides critical values of the tests for the 1%, 2.5%, 5% and 10% levels of significance. Note, however, that most econometric packages will report critical values and $p$ values for the Dickey-Fuller test regressions based on the

Table 5.2
The 1%, 2.5%, 5% and 10% critical values of the ADF test for Models 1, 2 and 3 for various sample sizes.

<table>
<thead>
<tr>
<th></th>
<th>T</th>
<th>1%</th>
<th>2.5%</th>
<th>5%</th>
<th>10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>25</td>
<td>-2.66</td>
<td>-2.26</td>
<td>-1.95</td>
<td>-1.60</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-2.62</td>
<td>-2.25</td>
<td>-1.95</td>
<td>-1.61</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-2.60</td>
<td>-2.24</td>
<td>-1.95</td>
<td>-1.61</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-2.58</td>
<td>-2.23</td>
<td>-1.95</td>
<td>-1.61</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-2.58</td>
<td>-2.23</td>
<td>-1.95</td>
<td>-1.61</td>
</tr>
<tr>
<td></td>
<td>∞</td>
<td>-2.58</td>
<td>-2.23</td>
<td>-1.95</td>
<td>-1.61</td>
</tr>
<tr>
<td>Model 2</td>
<td>25</td>
<td>-3.75</td>
<td>-3.33</td>
<td>-3.00</td>
<td>-2.62</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-3.58</td>
<td>-3.22</td>
<td>-2.93</td>
<td>-2.60</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-3.51</td>
<td>-3.17</td>
<td>-2.89</td>
<td>-2.58</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>-3.46</td>
<td>-3.14</td>
<td>-2.88</td>
<td>-2.57</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>-3.44</td>
<td>-3.13</td>
<td>-2.87</td>
<td>-2.57</td>
</tr>
<tr>
<td></td>
<td>∞</td>
<td>-3.43</td>
<td>-3.12</td>
<td>-2.86</td>
<td>-2.57</td>
</tr>
<tr>
<td>Model 3</td>
<td>25</td>
<td>-4.38</td>
<td>-3.95</td>
<td>-3.60</td>
<td>-3.24</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-4.15</td>
<td>-3.80</td>
<td>-3.50</td>
<td>-3.18</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-4.04</td>
<td>-3.73</td>
<td>-3.45</td>
<td>-3.15</td>
</tr>
<tr>
<td></td>
<td>∞</td>
<td>-3.96</td>
<td>-3.66</td>
<td>-3.41</td>
<td>-3.12</td>
</tr>
</tbody>
</table>

As an empirical illustration, the monthly United States zero coupon bond data are used to estimate Model 2 and Model 3, and p values for the tests are those derived from MacKinnon (1994). The p value for the Model 2 unit root test statistic (−2.120) is found to be 0.237; and because 0.237 > 0.05 the null hypothesis of nonstationarity cannot be rejected at the 5% level of significance. This outcome provides evidence that the interest rate is stochastically nonstationary and its generating mechanism has a unit root. For Model 3, the p value of the test statistic (−3.172) is found to be 0.091; because 0.091 > 0.05, the null hypothesis of a unit root is still not rejected at the 5% level of significance. This result is qualitatively the same as the unit root test based on Model 2, although there is a large reduction in the p-value from 0.237 in the case of Model 2 to 0.091 in Model 3.
5.3.2 The Augmented Dickey-Fuller (ADF) Test

In estimating any of the regression specifications given in equation (5.14), there is a real possibility that the disturbance term will exhibit autocorrelation. One reason for the presence of autocorrelation in the residuals is that many financial series interact with each other over time, which can induce joint dependence and serial dependence. Because the test regression specifications in (5.14), or their modifications given in (5.15), are univariate equations the effects of these interactions are ignored. Adjustments to the specifications to account for these interactions are therefore desirable in practical work.

One common solution to correct for induced autocorrelation is to proceed as suggested in Chapter 4 and include lags of the dependent variable \( \Delta y_t \) in the test regressions (5.14). Additional lagged variables help to remove this autocorrelation. With adjustments for the extra lagged variables, the equations take the augmented form

\[
\text{Model 1:} \quad \Delta y_t = \beta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + v_t, \\
\text{Model 2:} \quad \Delta y_t = \alpha + \beta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + v_t, \\
\text{Model 3:} \quad \Delta y_t = \delta_t + \beta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + v_t. 
\] (5.15)

In practice, the lag length \( p \) in these specifications is an unknown parameter and may be chosen by model selection methods, such as those outlined in Chapter 4, to ensure that the disturbances \( u_t \) do not exhibit autocorrelation. The unit root test procedures remain the same and involve the use of the same \( t \) statistic for testing \( \beta = 0 \) after taking into account the new specification of the various models given in (5.15).

The inclusion of lagged values of (differences in) the dependent variable represents an augmentation of the regression equation. Accordingly, the test is commonly referred to as the Augmented Dickey-Fuller (ADF) test, which was explored in Said and Dickey (1984). Setting \( p = 0 \) in any version of the test regressions in (5.15) gives the simple Dickey-Fuller test discussed earlier. Importantly, the asymptotic distributions of the ADF statistics in these augmented regressions are identical to the corresponding unit root asymptotic distributions in the original specifications (5.14). Hence, the same critical values as given in Table 5.2 may be used in conducting tests of the unit root hypothesis \( \beta = 0 \).

For example, using Model 2 in (5.15) to construct the ADF test with \( p = 2 \) lags for the United States zero coupon 2-month bond yield, the estimated regression equation is

\[
\Delta y_t = 0.092 - 0.017 y_{t-1} + 0.117 \Delta y_{t-1} - 0.080 \Delta y_{t-2} + \hat{v}_t. 
\]
The value of the ADF test statistic is
\[ t_\beta = \frac{\hat{\beta} - 0}{\text{se}(\hat{\beta})} = -0.017 - 0 \times 0.008 = -2.157. \]

Using the Dickey-Fuller limit distribution the \( p \)-value is 0.223. Since 0.223 > 0.05 the null hypothesis is not rejected at the 5% level of significance. This result is qualitatively the same as the earlier Dickey-Fuller test with \( p = 0 \) lags.

If \( p \) is chosen to be too small, then substantial autocorrelation may remain in the disturbance term in (5.15) and this will result in distorted statistical inference because the asymptotic distribution under the null hypothesis no longer applies in the presence of autocorrelation. However, including an excessive number of lags has an adverse effect on the ability of the test to reject the null hypothesis if it is false, referred to as a loss of power, because the extra lags serve to absorb variation in the data and tend to reduce its ability to discriminate.

To select the lag length \( p \) used in the ADF test, a common approach is to base the choice on information criteria such as those considered in Chapter 4. Two commonly used criteria are the Akaike Information criterion (AIC) and the Schwarz information criteria (SIC), otherwise known as the Bayesian information criterion (BIC). A lag-length selection procedure that is known to have good properties in unit root testing is the modified Akaike information criterion (MAIC) method proposed by Ng and Perron (2001). The lag length is chosen to satisfy

\[
\hat{p} = \arg \min_p \text{MAIC}(p) = \log(\hat{\sigma}_v^2) + \frac{2(\tau_p + p)}{T - p_{\text{max}}}, \tag{5.16}
\]

in which
\[ \tau_p = \frac{\hat{\beta}_2^2}{\hat{\sigma}_v^2} \sum_{t=p_{\text{max}}+1}^T \hat{u}_t^2, \]
and the maximum lag length is chosen as \( p_{\text{max}} = \text{int}[12(T/100)^{1/4}] \) where \( \text{int}[\cdot] \) signifies the integer part of the argument. In computing \( \hat{p} \), it is important that the sample over which the computations are performed is held constant.

5.4 Beyond the Simple Unit Root Framework

There is now a vast literature on unit root testing, much of which dealing with extensions of the framework just described to admit a far wider class of nonstationary time series. These extensions are particularly important in
financial applications because they allow for typical characteristics of financial data. A number of these developments are now standard in econometric software packages and are considered briefly below. Further extensions are considered in Chapter 11 where unit roots are applied to nonstationary panel data models.

5.4.1 Structural Breaks

The major form of nonstationarity considered so far is that of a stochastic trend which is generated by the presence of a unit root. There are other forms that nonstationarity in a time series may take and one of these occurs when the data are known to be disrupted by a permanent structural change. Such a break might arise from institutional changes, new policy implementations, or external shocks that have a permanent effect. The simplest structural break of this type involves a shift in the intercept of a time series, which delivers a level break or ‘crash’ in the event that the break is negative. Another type of structural break occurs when there is a shift in the slope as well as the intercept of a linear time trend. Breaks of this kind were discussed in the context of the use of dummy variables in Chapter 3.

When the timing of such structural breaks is known, it is straightforward to accommodate such shifts in the regression model. For instance, a simple mechanism for dealing with a level shift is to include a dummy variable in (5.15) to capture the structural break through the specification

\[
\Delta y_t = \beta y_{t-1} + \alpha + \delta t + \sum_{i=1}^p \phi_i \Delta y_{t-i} + \gamma \text{LBREAK}_t + v_t, \quad (5.17)
\]

where the structural break dummy variable is defined as

\[
\text{LBREAK}_t = \begin{cases} 
0 & : t \leq \tau \\
1 & : t > \tau,
\end{cases} \quad (5.18)
\]

and \(\tau\) is the observation (assumed to be known) where the break occurs.

Unit root tests are now constructed and performed just as before by testing the hypothesis \(\beta = 0\) in (5.17). However, because the regression equation (5.17) has changed through the inclusion of the covariate \(\text{LBREAK}_t\), the distribution of the ADF statistic under the null also changes to accommodate the presence of this covariate. Accordingly, the \(p\) values change from those of a statistic computed without these covariates and they now become a function of the timing of the structural break \(\tau\). These changes in the distribution theory mean that different tables of critical values must be employed when conducting a unit root test in the presence of breaks.

In a similar way, a companion variable can be introduced to capture a possi-
5.4. BEYOND THE SIMPLE UNIT ROOT FRAMEWORK

An additional shift in the time trend slope, as in the following specification

\[ \Delta y_t = \beta y_{t-1} + \alpha + \delta t + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + (\gamma a + \gamma d t) \text{LBREAK}_t + v_t. \]  

(5.19)

The structural break dummy variable \( \text{LBREAK}_t \) in (5.19) is the same as in (5.18) but now affects both the level and the time trend. Again, unit root tests are conducted in precisely the same manner as earlier. All that changes is the critical value used in the test. Tabulations of critical values for unit roots tests of the above type against alternative hypotheses that involve a structural break that occurs at some exogenously given date are calculated by simulation.

5.4.2 Generalised Least Squares Detrending

Consider the following model

\[ y_t = \alpha + \delta t + u_t, \]  

(5.20)

\[ u_t = \phi u_{t-1} + v_t, \]  

(5.21)

in which \( u_t \) is a disturbance term with zero mean and constant variance \( \sigma^2 \).

This is the fundamental equation from which Model 3 of the Dickey-Fuller test is derived. If the aim is still to test for a unit root in \( y_t \) the null and alternative hypotheses are

\[ H_0 : \phi = 1 \quad \text{[Nonstationary]} \]

\[ H_1 : \phi < 1. \quad \text{[Stationary]} \]  

(5.22)

Instead of proceeding in the manner described previously and using Model 3 in either (5.14) or (5.15), an alternative approach is to use a two-step procedure.

**Step 1: Detrending**

Estimate the parameters of equation (5.20) by ordinary least squares and then construct a detrended version of \( y_t \) given by

\[ y_t' = y_t - \hat{\alpha} - \hat{\delta} t. \]

**Step 2: Testing**

Test for a unit root using the deterministically detrended data, \( y_t' \), from the first step, using the Dickey-Fuller or augmented Dickey-Fuller test. Model 1 in (5.14) will be the appropriate model to use because, by construction, \( y_t' \) will have zero mean and no deterministic trend.
This procedure is equivalent to the single-step approach based on Model 3 in (5.14).

Elliott, Rothenberg and Stock (1996) suggested an alternative approach to detrending the data prior to testing for unit root. Their approach aims to address an important shortcoming of the Dickey-Fuller approach that the tests may have low power. The modified detrending approach is based on the premise that the test is more likely to reject the null hypothesis of a unit root if under the alternative hypothesis the detrending process takes into account that the process may have a root that is in the region of unity (more specifically, local to unity in a manner dependent on the sample size). The modified detrending step proceeds as follows. Define the constant $\phi^*$ in specific local to unity form as

$$\phi^* = 1 + \tau/T$$

where $\tau = \begin{cases} 
-7 & \text{[Constant ($\alpha \neq 0, \delta = 0$)]} \\
-13.5 & \text{[Trend ($\alpha \neq 0, \delta \neq 0$)]} 
\end{cases}$

and use it to construct the following variables

$$y^*_t = y_t - \phi^*y_{t-1},$$
$$\alpha^* = 1 - \phi^*,$$
$$t^* = t - \phi^*(t-1).$$

The starting values for each of these variables at $t = 1$ are $y^*_1 = y_1$ and $\alpha^*_1 = 1$ and $t^*_1 = 1$, respectively. The starting values are important because if $\tau = -T$ then this differencing procedure has no effect. If, on the other hand, $\tau = 0$ then the procedure reverts to a simple first difference. This kind of detrending is commonly referred to as generalised least squares detrending, but is also known as quasi-differencing and partial generalised least squares (Phillips and Lee, 1995).

The choice of the value of the constant $\tau$ in this detrending process is determined so that the test reaches the envelope of maximum power under certain conditions when optimal power is around 50%. In this sense the test is considered to be point optimal for local departures from unity with autoregressive coefficient $\phi = 1 + c/T$ and localising coefficient $c = \tau$. For example, based on a sample size of $T = 200$, $\phi^* = 1 + \tau = 1 - 7/200 = 0.9650$ for a regression with only a constant and $0.9325$ for a regression with a constant and a time trend.

Using the newly defined variables in (5.23), run the regression

$$y^*_t = \pi_0 \alpha^* + \pi_1 t^* + u^*_t,$$

in which $u^*_t$ is a composite disturbance term. Once the ordinary least squares estimates $\widehat{\pi}_0$ and $\widehat{\pi}_1$ are available, detrended data

$$\widehat{u}^*_t = y^*_t - \widehat{\pi}_0 \alpha^* - \widehat{\pi}_1 t^*,$$
can be constructed. The detrended data from (5.25) can now be tested for a unit root. If Model 1 of the Dickey-Fuller framework is used then the test is referred to as the GLS-DF test. Note, however, that because the detrended data depend on the value of $c$ the critical values are different to the Dickey-Fuller critical values which rely on simple detrending.

This modified approach to unit root testing relies on the assumption that departures from unit roots are uniformly of the local to unity form $\phi = 1 + c/T$. Many realistic departures (including structural breaks) are not of this simple form and, in such cases, GLS tests suffer power losses, are no longer point optimal tests, and do not necessarily dominate standard unit root tests (Bykhovskaya and Phillips, 2018).

### 5.4.3 Nonparametric Adjustment for Autocorrelation

Phillips and Perron (1988) proposed an alternative method for adjusting the Dickey-Fuller test for autocorrelation of general form and some forms of heterogeneity. The adjustment is nonparametric rather than parametric and therefore produces a unit root test of wide applicability. The test is based on estimating the Dickey-Fuller regression equation, either (5.10) or (5.13), by ordinary least squares but the test statistic uses a nonparametric approach to correct for the autocorrelation. The Phillips-Perron statistic is

$$
\tilde{t}_\beta = t_\beta \left( \frac{\hat{\gamma}_0}{\hat{f}_0} \right)^{1/2} - \frac{T(\hat{f}_0 - \hat{\gamma}_0)se(\hat{\beta})}{2\hat{f}_0^{1/2}s},
$$

where $t_\beta$ is the ADF statistic, $s$ is the standard error of the Dickey-Fuller test regression, and $\hat{f}_0$ is known as the long-run variance. The long-run variance is computed as

$$
\hat{f}_0 = \hat{\gamma}_0 + 2 \sum_{j=1}^{p} \left( 1 - \frac{j}{p} \right) \hat{\gamma}_j,
$$

where $p$ is the length of the lag, and $\hat{\gamma}_j$ is the $j^{th}$ estimated autocovariance function of the ordinary least squares residuals obtained from estimating either (5.10) or (5.13)

$$
\hat{\gamma}_j = \frac{1}{T-j} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}.
$$

The critical values are the same as the Dickey-Fuller critical values when the sample size is large, which makes this general test very easy to apply in practical work.
5.4.4 Unit Root Testing with a Null of Stationarity

The Dickey-Fuller testing framework, including the GLS detrending and Phillips-Perron variants, are designed for testing the null hypothesis that a time series $y_t$ is nonstationary or $I(1)$. There is also a popular test commonly known as the KPSS test, after Kwiatkowski, Phillips, Schmidt and Shin (1992) that is often reported in the empirical literature which has a null hypothesis of stationarity or $I(0)$. Consider the regression model

$$y_t = \alpha + \delta t + w_t + u_t,$$

where $w_t$ is given by

$$w_t = w_{t-1} + v_t, \quad v_t \sim iid \ N(0, \sigma_v^2).$$

The null hypothesis that $y_t$ is a stationary $I(0)$ process is tested in terms of the null hypothesis $H_0 : \sigma_v^2 = 0$ in which case $w_t$ is simply a constant equal to zero. Define $\{\hat{u}_1, \cdots, \hat{u}_T\}$ as the ordinary least squares residuals from a regression of $y_t$ on a constant and a deterministic trend. Now define the standardised test statistic

$$S = \frac{1}{T^2 f_0} \sum_{t=1}^{T} (\sum_{j=1}^{T} \hat{u}_j)^2,$$

in which $f_0$ is a consistent estimator of the long-run variance of $u_t$, as defined in equation (5.27).

5.5 Asset Price Bubbles

During the 1990s, led by Dot-Com stocks and the internet sector, the United States stock market experienced a spectacular rise in all major indices, especially the NASDAQ index. Figure 5.5 plots the monthly NASDAQ index, expressed in real terms, for the period February 1973 to January 2009. The series grows fairly steadily until the early 1990s and then begins to surge. The steep upward movement in the series continues until the late 1990s as investment in Dot-Com stocks grows in popularity. Early in the year 2000 the Index drops abruptly and then continues to fall to the mid-1990s level. In summary, over the decade of the 1990s, the NASDAQ index rose to its historical high on 10 March 2000 and then collapsed, in the process creating and destroying some US$8 trillion dollars of shareholder wealth.

Concomitant with the striking rise in stock market indices during this period there was much popular talk among economists about the effects of the internet and computing technology on productivity and the emergence of a new economy associated with these changes. The information revolution, it was argued, may have provided a new fundamental driver that explained the surge in stock prices. What actually caused this and other unusual surges
5.5. ASSET PRICE BUBBLES

and subsequent falls in prices, whether there were bubbles and, if so, whether they were rational or behavioural are among the most actively debated issues in macroeconomics and finance.

Figure 5.5: The monthly NASDAQ index expressed in real terms for the period February 1973 to January 2009.

A recent line of research has developed empirical tests for bubbles and rational exuberance in financial asset markets, with applications to phenomena such as the Dot-Com bubble, the US property market bubble, and the 2008 global financial crisis (Phillips and Yu, 2011; Phillips, Wu and Yu, 2011). This development provides an interesting new variant in the field of unit root testing. Instead of concentrating on performing a test of a unit root against the alternative of stationarity (which employs a one-sided test where the critical region is defined in the left-hand tail of the distribution of the test statistic), financial time series may be tested for explosive behaviour by testing against the alternative of an explosive root, leading to a right-sided unit root test where the critical value of the test lies in the right tail of the distribution. Such tests are appropriate for asset prices exhibiting exuberance or bubble-like behaviour. The null hypothesis of interest is still $\rho = 1$ but the alternative hypothesis is now $\rho > 1$ in (5.6), or

$$
H_0 : \rho = 1 \quad \text{[Variable is nonstationary, no price bubble]} \\
H_1 : \rho > 1. \quad \text{[Variable is explosive, price bubble]} 
$$

(5.29)

To motivate the presence of a price bubble, consider the following model

$$
P_t (1 + R) = E_t [P_{t+1} + D_{t+1}] ,
$$

(5.30)
where \( P_t \) is the price of an asset, \( R \) is the risk-free rate of interest assumed to be constant for simplicity, \( D_t \) is the dividend payment and \( E_t[\cdot] \) is the conditional expectations operator that conditions on information up to time \( t \). This equation highlights two types of investment strategies. The first is given by the left hand-side which involves investing in a risk-free asset at time \( t \) yielding a payoff of \( P_t (1 + R) \) in the next period. Alternatively, the right-hand side shows that by holding the asset the investor earns the capital gain from owning an asset with a potentially higher price the next period plus a dividend payment. In equilibrium there are no arbitrage opportunities so the two types of investment are equal to each other.

Now divide both sides of equation (5.30) by \((1 + R)\) to yield
\[
P_t = \beta E_t [P_{t+1} + D_{t+1}],
\]
(5.31)
where \( \beta = (1 + R)^{-1} \) is the discount factor. Writing this expression at \( t + 1 \)
\[
P_{t+1} = \beta E_t [P_{t+2} + D_{t+2}],
\]
(5.32)
means that it can be used to substitute out \( P_{t+1} \) in (5.31)
\[
P_t = \beta E_t [\beta E_t [P_{t+2} + D_{t+2}] + D_{t+1}] = \beta E_t [D_{t+1}] + \beta^2 E_t [D_{t+2}] + \beta^2 E_t [P_{t+2}].
\]
Repeating this approach \( N - 1 \) times gives the price of the asset in terms of two components
\[
P_t = \sum_{j=1}^{N} \beta^j E_t [D_{t+j}] + \beta^N E_t [P_{t+N}]
\]
(5.33)
\[
= \sum_{j=1}^{N} \beta^j E_t [D_{t+j}] + B_t,
\]
where
\[
B_t = \beta^N E_t [P_{t+N}].
\]
The first term on the right-hand side of (5.33) is the standard present value of an asset whereby the price of an asset equals the discounted present value stream of expected dividends and the second term is a bubble component, Diba and Grossman (1988) argue that this bubble component has an explosive property given by
\[
E_t [B_{t+1}] = (1 + R)B_t.
\]
In the absence of bubbles \((B_t = 0)\), the degree of nonstationarity of the asset price is controlled by the character of the dividend series, but asset prices will be explosive in the presence of bubbles because \(1 + R > 1\). The implied conditional expectation at time \( t \) of the value of the process \( h \) periods ahead is
\[
E_t [B_{t+h}] = (1 + R)^h B_t.
\]
It follows that $E_t [B_t+h]$ grows exponentially with the horizon $h$. The process $B_t$ is known as the rational bubble component of the solution (5.33) and will always be present in the solution unless the initial value $B_0 = 0$.

As indicated earlier, one way of testing for the presence of a bubble process like $B_t$ in asset prices is to apply a right-sided unit root test to the price process itself. Interestingly enough, if that convention were followed and the ADF test were applied to the full sample (February 1973 to January 2009), the unit root test would not reject the null hypothesis $H_0: \rho = 1$ in favour of the right-tailed alternative $H_1: \rho > 1$ at the 5% level of significance. This outcome would lead to the conclusion that there is no significant evidence of exuberance in the behaviour of the NASDAQ index over the sample period. This conclusion would sit comfortably with a consensus view that was held before the NASDAQ experience that there is little empirical evidence to support the hypothesis of explosive behaviour in stock prices (see, for example, Campbell, Lo and MacKinlay, 1997, p260).

On the other hand, Evans (1991) argued that explosive behaviour in prices is only temporary in the sense that bubbles eventually collapse. Consequently, observed trajectories of asset prices may appear more like an I(1) or even a stationary series rather than an explosive one even when bubbles occur in the trajectory, thereby confounding empirical evidence about the existence of bubbles in financial data. Using simulations, Evans demonstrated that standard unit root tests have difficulties in detecting such periodically collapsing bubbles.

To address the lack of power of the full sample-based unit root test in detecting periodically collapsing bubbles, PWY (2011) suggest implementing recursively a unit root test based on expanding windows of observations, starting with $T_0 = \lfloor T r_0 \rfloor$ observations in the first regression and ending with $T$ observations in the final regression, where $T$ is the full sample size and $r_0 \in (0, 1)$ is the sample fraction of $T$ used for initialising the test recursion. The test statistic is the maximum of the $t$ statistics obtained in this way over the expanding windows of data. The asymptotic distribution of the test statistic is then the supremum of the unit root distributions taken over this range of values. Critical values for the recursive test are obtained by simulation. The test then provides a test of explosive behaviour over the sample period $[T_0, T]$.

An important outcome of this testing framework is that it also delivers a date-stamping procedure for the origination and termination of bubbles. By matching the recursive $t$ statistics with the path of the right-tailed critical values an estimate of the origination of bubble behaviour in the data is obtained by determining the first observation for which the test statistic crosses the critical value path. This is known as the first crossing time principle. Similarly, an estimate of the termination of the bubble is obtained by noting the observation for which the recursive test statistic crosses back over and falls below the critical value path. This use of recursive unit root testing provides a valuable approach to the detection and dating of bubbles in financial data. The method
is now employed by central banks, financial institutions, and regulators to assess in real time prevailing financial market conditions.

Figure 5.6: The Phillips, Wu and Yu testing procedure to test for price bubbles in the monthly NASDAQ index expressed in real terms for the period February 1973 to January 2009 by means of recursive expanding window Augmented Dickey Fuller tests with 1 lag. The startup sample is 39 observations from February 1973 to April 1976. The approximate 5% critical value is shown by the dashed line in the Figure.

Figure 5.6 is a plot of the forward recursive ADF statistic with 1 lag computed from forward recursive regressions by fixing the start of the sample period and progressively increasing the sample size observation by observation until the entire sample is used. The startup sample is 39 observations. The NASDAQ shows no evidence of rational exuberance prior to June 1995 as the ADF statistics lie below the dashed line. In July 1995, the test detects the presence of explosive behaviour (which can be interpreted as exuberance) in stock prices ($\rho > 1$), which is signified by a significant test statistic. The supporting evidence for the presence of a bubble tends to become progressively stronger in terms of the test statistic’s deviation from the critical value path and it reaches a peak in February 2000. The bubble continues until February 2001 and by March 2001 the bubble appears to have dissipated. Interestingly, the first occurrence of the bubble is July 1995, which is more than a year before the remark by Alan Greenspan, the then Chairman of the Federal Reserve Board, on 5 December 1996, which coined the phrase irrational exuberance to characterise irrationally optimistic herding behaviour in stock markets.

To check the robustness of these findings, Figure 5.7 plots the ADF statistic
5.5. ASSET PRICE BUBBLES

The Phillips, Shi and Yu testing procedure to test for price bubbles in the monthly NASDAQ index expressed in real terms for the period February 1973 to January 2009 by means of rolling window ADF tests with 1 lag. The size of the window is set to 77 observations so that the starting sample is February 1973 to June 1979. The approximate 5% critical value is shown by the dashed line in the Figure.

Figure 5.7: The Phillips, Shi and Yu testing procedure to test for price bubbles in the monthly NASDAQ index expressed in real terms for the period February 1973 to January 2009 by means of rolling window ADF tests with 1 lag. The size of the window is set to 77 observations so that the starting sample is February 1973 to June 1979. The approximate 5% critical value is shown by the dashed line in the Figure.

with 1 lag for a series of rolling window regressions. Each regression is based on a subsample of size $T = 77$ with the first sample period from February 1973 to June 1979. The fixed window is then rolled forward one observation at a time. The general pattern to emerge is completely consistent with the results reported in Figure 5.6. Tests for exuberance based on these rolling window regressions and recursively evolving versions of such regressions are developed in Phillips, Shi and Yu (2015b, 2015a).

Given that the model used to produce these results has a reduced form structure, the findings do not deliver a causal explanation for the exuberance of the 1990s in Dot-Com stocks. Several possibilities exist, including the presence of a rational bubble, herding behaviour, or explosive effects on economic fundamentals arising from time variation in discount rates. Identification of explicit economic source(s) of this observed price behaviour will involve more explicit formulation of structural models of behaviour that enables identification of the driving force(s).

What the recursive testing methodology does provide is clear support for the presence of a mildly explosive propagating mechanism in the NASDAQ index and estimates of the origination and termination of this phenomena in the
data. A further advantage of the recursive methodology is that it can be used in real time to make ongoing assessments of the state of the stock market. The methods can be applied in this way to study recent phenomena in many different markets such as real estate, commodity, foreign exchange, and equity markets, all of which have attracted considerable attention. They are now being used by regulators and central banks to assist in monitoring the state of financial markets.

5.6 Exercises

The data required for the exercises are available for download as EViews workfiles (*.wfl), Stata datafiles (*.dta), comma delimited text files (*.csv) and as Excel spreadsheets (*.xlsx).

1. Commodity Price Data

<table>
<thead>
<tr>
<th>commodity.*</th>
</tr>
</thead>
</table>

The data comprise monthly observations on a number of United States traded commodity price indices for the period January 1957 to November 2003.

(a) Use the commodity prices to construct the following transformed series; the natural logarithm of commodity prices, the first difference of commodity prices and log returns of commodity prices. Plot the series and discuss the stationarity properties of each series.

(b) Use the following unit root tests to test for stationarity of the natural logarithm of commodity prices.
   i. Dickey-Fuller test with a constant and no time trend.
   ii. Augmented Dickey-Fuller test with a constant and no time trend and \( p = 2 \) lags.
   iii. Phillips-Perron test with a constant and no time trend and \( p = 2 \) lags.
   iv. Repeat parts (ii) and (iii) where the lag length for the ADF and PP tests is based on a data driven selection of the optimal lag length.

(c) Repeat part (b) for the first differences and log returns of commodity prices.

(d) Now construct the first difference of log returns of commodity prices. Repeat part (b). Are you able to reach a conclusion about the appropriate level of differencing, \( d \), required to achieve stationarity for the logarithm of commodity prices, assuming the \( d \) takes an integer value.
2. Equity Market Data

The data are monthly observations on United States equity prices and dividends for the period January 1871 to September 2016.

(a) Use the equity price series to construct the following transformed series; the natural logarithm of equity prices, the first difference of equity prices and log returns of equity prices. Plot the series and discuss the stationarity properties of each series.

(b) Construct similarly transformed series for dividend payments and discuss the stationarity properties of each series.

(c) Construct similarly transformed series for earnings and discuss the stationarity properties of each series.

(d) Use the following unit root tests to test for stationarity of the natural logarithms of prices, dividends and earnings.
   i. Dickey-Fuller test with a constant and no time trend.
   ii. Augmented Dickey-Fuller test with a constant and no time trend and \( p = 1 \) lag.
   iii. Phillips-Perron test with a constant and no time trend and \( p = 1 \) lag.
   iv. Repeat parts (ii) and (iii) where the lag length for the ADF and PP tests is based on a data driven selection of the optimal lag length.

(e) Repeat part (b) for the first differences and log returns of equity prices, dividends and earnings.

3. Unit Root Tests of Bond Market Data

The data are monthly observations from December 1946 to February 1987 on United States zero coupon bond yields for maturities ranging from 2 months to 9 months.

(a) Use the following unit root tests to determine the stationarity properties of each yield.
   i. Dickey-Fuller test with a constant and no time trend.
   ii. Augmented Dickey-Fuller test with a constant and no time trend, and \( p = 2 \) lags.
iii. Phillips-Perron test with a constant and no time trend, and \( p = 2 \) lags.
iv. Repeat parts (ii) and (iii) where the lag length for the ADF and PP tests is based on a data driven selection of the optimal lag length.

(b) Now construct the spreads between the longer-term maturities and the 1-month yield. Repeat (a) and hence test the stationarity properties of the spreads.

4. Fisher Hypothesis

The data are quarterly observations on United States inflation and interest rates from 1954:Q3 to 2008:Q4. The Fisher hypothesis states that the nominal interest rate fully reflects the long-run movements in the inflation rate.

(a) Construct the percentage annualised inflation rate, \( \pi_t \).
(b) Plot the nominal interest rate and inflation rate.
(c) Perform unit root tests to determine the level of integration of the nominal interest rate and inflation. In performing the unit root tests, test the sensitivity of the results by using a model with a constant and no time trend, and a model with a constant and a time trend. Let the lags be determined by the automatic lag length selection procedure. Discuss the results in terms of the level of integration of each series.
(d) Compute the real interest rate as
\[
\Delta r_t = i_t - \pi_t.
\]
Test the real interest rate \( r_t \) for stationarity using a model with a constant but no time trend based on a data driven selection of optimal lag length. Does the Fisher hypothesis hold? Discuss.

5. Price Bubbles in the Share Market

The data are monthly observations of the Nasdaq price index and the dividends for the period January 1960 to December 1989. The present value model implies that the fundamental driver of the share price \( P_t \) is the dividend payment \( D_t \). A bubble occurs when the actual share price persistently deviates from fundamentals.
(a) Create the log price-dividend ratio, \( p_t - d_t \), where as customary lower case variables refer to logarithms. Use right-tailed unit root tests applied to the entire sample to determine whether or not a bubble exists. Take the minimum window size to be 39 observations.

(b) Use the forward recursive right-tailed Dickey-Fuller test procedure to test for a bubble in the log price-dividend ratio.

(c) Use a rolling window right-tailed Dickey-Fuller test procedure to test for a bubble in the log-price dividend ratio. Discuss your results.
Chapter 6

Cointegration

As discussed in Chapter 5, a single nonstationary time series can often be rendered stationary by differencing. In a multivariate context, the same differencing technique may be applied to each component to achieve stationarity. But an alternative method may also be available to remove nonstationarity. When there is more than one nonstationary time series and the series tend to move together over time, it is often possible to form linear combinations of the nonstationary series that produce a stationary series. This ability to generate stationary time series through carefully chosen linear combinations of nonstationary time series is known as cointegration. The weights that together constitute this special linear combination of the original series are known as cointegrating coefficients.

The idea that nonstationary time series may be reduced to stationary time series by means of well-defined regression specifications has long been implicit in much empirical research in economics as well as the adjustment mechanisms that underpinned the form of many models formulated in economic theory for aggregate economic data, such as trade cycle and cyclical growth models in macroeconomics. Bergstrom (1967) provides an extensive discussion of such models and many references. In econometrics, the central contribution to this subject was made by Engle and Granger (1987) and acknowledged in the award of the 2004 Nobel Prize in Economic Science.

The concept of cointegration provides a useful basis for interpreting many financial models in terms of meaningful long-run relationships. Such relationships are often obtained as equilibrium path solutions of inter-temporal decision making problems that face economic agents or in linkages that explain asset prices in terms of expected returns. The empirical relevance of such theories can then be investigated in applied work by examining evidence in support of cointegration. Having uncovered long-run relationships between financial variables by establishing evidence of cointegration, the transient
or short-run properties of the variables may be separately modelled or even modelled simultaneously by building transient responses directly into the specification of the equations that define the long-run relationships. The latter type of model is useful because it can sometimes be interpreted in terms of the adjustment mechanisms present in some underlying financial theory. The approach has a practical advantage because the econometric specification, which is known as a vector error-correction model (VECM), is simply a restricted form of the commonly used vector autoregressive (VAR) model considered in Chapter 4.

The existence of cointegration between nonstationary time series has important theoretical, statistical and dynamic implications. Among these are the following.

(i) A number of theoretical models in finance can be couched within a cointegrating framework.

(ii) Estimates of the parameters in the cointegrating equations, including ordinary least squares estimates, converge to their population values at a rate faster than is the case for stationary variables, a property known as super-consistency.

(iii) Special methods of estimation have been developed to provide asymptotically efficient estimates, to address bias and serial dependence, and to facilitate inference. These methods are easy to use, some involve simple modifications of least squares procedures, and many are now available in standard software packages.

(iv) Modelling a system of cointegrated variables allows for the joint specification of long-run and short-run dynamics of financial variables in terms of the VECM.

6.1 The Present Value Model and Cointegration

An important property of asset prices identified in both Chapters 2 and 5 is that they exhibit stochastic trends. This is indeed the case for the United States as seen in Figure 6.1 which shows that the logarithms of monthly equity prices, \( p_t = \log P_t \), dividend payments, \( d_t = \log D_t \), and earnings \( e_t = \log E_t \) all exhibit strong (largely positive) trends over the period January 1871 to September 2016. These series are indicative of many financial time series that exhibit trending, nonstationary characteristics that are well-represented as integrated \( I(1) \) processes, as discussed in Chapter 5.
6.1. THE PRESENT VALUE MODEL AND COINTEGRATION

6.1.1 Equilibrium Relationships

Even though the pattern of behaviour of the series in Figure 6.1 indicates that equity prices and dividends are trending $I(1)$ series, the present value model of equity prices provides a theoretical link between equity prices and dividends that is strongly suggestive of co-movement in the two series over time. From Chapter 2 this relationship is expressed as

\[ p_t = \beta_0 + \beta_d d_t + u_t, \]  

(6.1)

where $p_t$ is the log equity price, $d_t$ is the log dividend, $u_t$ is a disturbance term and $\beta_0$ and $\beta_d$ are unknown parameters. Both $p_t$ and $d_t$ are well-modelled as $I(1)$ processes. Yet the present value model indicates that the differences $p_t - \beta_0 - \beta_d d_t = u_t$ are simply transient shocks that do not disturb the nature of the relationship (6.1) over time. The linkage between $p_t$ and $d_t$ in (6.1) is therefore regarded as a long run (or permanent) relationship between trending $I(1)$ series and the residuals $u_t$ are viewed as transient shocks or $I(0)$. The linear combination of the $I(1)$ variables $p_t$ and $d_t$ which results in the $I(0)$ variable given by $u_t$ is known as a cointegrating relationship or simply as cointegration. The reduction of the trending $I(1)$ character of $p_t$ and $d_t$ to the transient $I(0)$ character of $u_t$ is the essential condition of cointegration. When these conditions hold, then (6.1) is known as a cointegrating equation and the parameters $\beta_0$ and $\beta_d$ are the cointegrating coefficients.

Figure 6.1: Time series plots of the logarithms of monthly United States equity prices, dividend payments and earnings per share for the period February 1871 to September 2016.
An alternative way of viewing the present value model in (6.1) as a cointegrating system is through the scatter diagram in Figure 6.2 for \( p_t \) and \( d_t \) using the same data presented in Figure 6.1. Superimposed on the scatter plot is an estimate of the present value model equation

\[
p_t = 3.1375 + 1.1957 d_t + \hat{u}_t, \tag{6.2}
\]

obtained by regressing \( p_t \) on a constant and \( d_t \). Even though \( p_t \) and \( d_t \) are both nonstationary, prices and dividends never deviate too far away from the line given by \( 3.1375 + 1.1957 \ d_t \), suggesting that this relationship captures an equilibrium path between these variables. If there were no cointegration between the variables the scatter diagram would have points much more evenly scattered about the two dimensional plane. In other words, cointegration has the effect of an attractor amongst nonstationary series, in the present case compressing equity prices and dividends close to a one dimensional relationship.

This interpretation of the scatter diagram also suggests that actual movements in \( p_t \) can be decomposed into a long-run component representing the (dynamic) equilibrium price as determined by dividends and a short-run, transient component that represents temporary deviations of \( p_t \) from its long-run path. This decomposition is expressed as

\[
\begin{align*}
\text{Actual path} & = \beta_0 + \beta_d d_t + u_t \\
\text{Long-run path} & \\
\text{Short-run deviations} & 
\end{align*}
\]

### 6.1.2 Equilibrium Dynamics

The scatter plot between equity prices and dividends in Figure 6.2 suggests that not only are the shocks arising from \( u_t \) transitory, but that there are forces continually pushing the system back towards the equilibrium path whenever there is a shock. These adjustment dynamics are highlighted in Figure 6.3 where the line containing the points ADC represents the equilibrium line of the present value model in equation (6.2) and pressure from points B and E above and below the line drive the system towards the equilibrium path.

To understand the equilibrium forces within the present value model consider the effects of a shock \( u_t \) at time \( t \), assuming initially that dividends \( d_t \) are unaffected by this shock. For a positive shock with \( u_t > 0 \), equities appear overvalued relative to the long-run level since the stock price \( p_t \) lies above its long-run equilibrium price determined by dividend flows (Point B in Figure 6.3). Similarly, with a negative shock \( u_t < 0 \), equities appear undervalued because \( p_t \) lies below its long-run equilibrium price (Point E in Figure 6.3).

Consider the effect of a positive shock that takes the share price above its equilibrium price at Point A. The disequilibrium induced by this price shock
in the system leads to forces of adjustment that operate to drive prices back towards the equilibrium path somewhere along the ADC line in Figure 6.3. There are three potential scenarios to consider.

(i) **Equity prices adjust**

For equilibrium to be re-established in this scenario equity prices need to decrease back to point A from point B without any change in dividends. Assuming that the fall in equity prices is proportional to the size of the shock $u_t = p_t - \beta_0 - \beta_d d_t$, the change in the equity price in the next period $p_{t+1} - p_t$, is represented by the adjustment equation

$$p_{t+1} - p_t = \delta_1 + \alpha_1 (p_t - \beta_0 - \beta_d d_t) + v_{1t+1},$$  \hspace{1cm} (6.3)

where $v_{1t+1}$ is a disturbance term capturing additional future movements in $p_t$ not arising from the shock $u_t$. Given that equity prices need to adjust downwards in this scenario to restore equilibrium, the adjustment parameter $\alpha_1$ satisfies the restriction $\alpha_1 < 0$. 

---

Figure 6.2: Scatter plots of the logarithms of monthly United States equity prices and dividends, in panel (a), and equity prices and earnings per share, in panel (b), for the period February 1871 to September 2016.
(ii) **Dividends adjust**

For equilibrium to be re-established in this scenario dividends need to increase from point B to point C while equity prices remain fixed. Adopting the assumption that the adjustment in dividends is also proportional to the size of the shock \( u_t = p_t - \beta_0 - \beta_d d_t \), the change in future dividends is represented by the adjustment equation

\[
\Delta d_{t+1} = \delta_2 + \alpha_2 (p_t - \beta_0 - \beta_d d_t) + v_{2t+1}, \tag{6.4}
\]

where \( v_{2t+1} \) is a disturbance term capturing additional future movements in \( d_t \) not arising from the initial shock \( u_t \). Given that dividends need to increase to restore equilibrium in this scenario, the adjustment parameter \( \alpha_2 \) satisfies the restriction \( \alpha_2 > 0 \).

(iii) **Equity prices and dividends adjust**

In this scenario both the equity price and dividend adjustment equations (6.3) and (6.4) operate with \( p_t \) decreasing and \( d_t \) increasing. The relative strength of the movements in equity prices and dividends is determined by the relative magnitudes of the adjustment parameters \( \alpha_1 \) and \( \alpha_2 \). If, for example, the movement is from point B to point D as shown in Figure 6.3 then both equity prices and dividends bear an equal share of the adjustment towards equilibrium.

As it is equity prices that adjust over time in the first scenario towards equilibrium this variable is classified as endogenous, whereas the dividend variable is treated as exogenous\(^1\) given that it does not adjust to restore equilibrium in response to the disequilibrium \( p_t - \beta_0 - \beta_d d_t \) and \( \alpha_2 = 0 \) accordingly.

---

\(^1\)The definition of exogeneity is defined more precisely later, taking into account certain refinements of the concept such as weak and strong exogeneity.
6.2 VECTOR ERROR CORRECTION MODELS

The opposite is the case for the second scenario where the dividend variable is the endogenous variable and the equity price is treated as exogenous, setting $\alpha_1 = 0$. For the third scenario equity prices and dividends are jointly determined and endogenous as both variables adjust the system towards equilibrium, and in this event $\alpha_1, \alpha_2 \neq 0$.

To gauge the relative strength of the adjustment parameters, $\alpha_1$ and $\alpha_2$, in restoring equilibrium in the present value model for the United States, equations (6.3) and (6.4) are estimated by ordinary least squares using the equity price and dividends data presented in Figure 6.1. Letting $\hat{u}_t$ represent the ordinary least squares residuals from (6.2), the parameter estimates of the price adjustment equation are computed by regressing $\Delta p_{t+1}$ on a constant and $\hat{u}_t$, with the estimated model given by

$$\Delta p_{t+1} = 0.0035 - 0.0011 \hat{u}_t + \hat{v}_{1t+1}. \quad (6.5)$$

Similarly, the parameter estimates of the dividend adjustment equation are computed by regressing $\Delta d_{t+1}$ on a constant and $\hat{u}_t$ resulting in the estimated model

$$\Delta d_{t+1} = 0.0029 + 0.0078 \hat{u}_t + \hat{v}_{2t+1}. \quad (6.6)$$

An alternative representation of the estimated adjustment equations in (6.5) and (6.6) is to use the estimated long-run equilibrium relationship in equation (6.2) to substitute out $\hat{u}_t$ so that

$$\Delta p_{t+1} = 0.0035 - 0.0011 (p_t - 3.1375 - 1.1957 d_t) + \hat{v}_{1t+1}, \quad (6.7)$$

$$\Delta d_{t+1} = 0.0029 + 0.0078 (p_t - 3.1375 - 1.1957 d_t) + \hat{v}_{2t+1}.$$\)

The signs of the adjustment parameter estimates are consistent with the dynamics presented in Figure 6.3 with equity prices decreasing ($\hat{\alpha}_1 < 0$) and dividends increasing ($\hat{\alpha}_2 > 0$) to restore equilibrium after a positive shock to equity prices. Moreover, since

$$|\hat{\alpha}_2| = 0.0078 > |\hat{\alpha}_1| = 0.0011,$$

dividends appear to be the stronger driving force in restoring equilibrium in the system after a shock to the equity price.

6.2 Vector Error Correction Models

The bivariate set of equations containing equity prices and dividends in (6.3) and (6.4) is known as a vector error correction model or VECM. To generalise this model it is convenient to re-express the present value cointegrating model as

$$\Delta y_{1t} = \delta_1 + \alpha_1 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1}) + v_{1t},$$

$$\Delta y_{2t} = \delta_2 + \alpha_2 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1}) + v_{2t}. \quad (6.8)$$
where \( y_{1t} \) and \( y_{2t} \) are two \( I(1) \) variables, and for convenience the time subscript \( t \) is redefined by replacing \( t + 1 \) by \( t \). The parameters \( \beta_0 \) and \( \beta_2 \) represent the cointegrating or long-run parameters with the long-run parameter on \( y_{1t} \), namely \( \beta_1 \) normalised to unity. The adjustment parameters \( \alpha_1 \) and \( \alpha_2 \) are known as the error correction parameters as they control the relative strengths of the adjustments in the dependent variables with respect to the (lagged) equilibrium error. These terms control the error-correction mechanism. Finally, the parameters \( \delta_1 \) and \( \delta_2 \) allow the time series \( y_{1t} \) and \( y_{2t} \) to have some deterministic drift. However, a model that is often used in empirical work imposes the restriction \( \delta_1 = \delta_2 = 0 \), which is appropriate if the two time series \( y_{1t} \) and \( y_{2t} \) exhibit random walk behaviour without evidence of any deterministic trends.

### 6.2.1 Extensions

VECM specifications such as (6.8) often include additional endogenous variables and more complex short-run dynamics that are captured using extra lags of the endogenous variables, or equivalently, changes in those variables. An important extension of the VECM in empirical work is to allow for multiple cointegrating equations. This extension is appropriate where it is known or hypothesised that there are several equilibrium relationships operating simultaneously among the \( I(1) \) variables. This multivariate case is explored at a later stage in the present chapter. Deterministic variables such as time trends can also be included in the same way that these variables appear in unit root models and tests. In multivariate settings, deterministic variables may also appear in the long-run cointegrating equation in which case the variables are considered to be deterministically as well as stochastically cointegrated.

### Additional \( I(1) \) Variables

The present value model is a bivariate model containing two \( I(1) \) variables with \( y_{1t} \) representing log equity prices and \( y_{2t} \) representing log dividend payments. This model is extended in Section 6.6 to a trivariate model that allows for earnings. In the models of the term structure of interest rates discussed in Section 6.7 the dimension of the system is governed by the number of yields with different maturities that are included in the model. Extending the framework from a bivariate model as in (6.8) to \( N \) variables that are all \( I(1) \) with a common scalar long-run relationship simply involves augmenting the two
VECM equations in (6.8) to \( N \) equations according to the following scheme

\[
\begin{align*}
\Delta y_{1t} &= \delta_1 + \alpha_1 u_{t-1} + v_{1t}, \\
\Delta y_{2t} &= \delta_2 + \alpha_2 u_{t-1} + v_{2t}, \\
&\vdots \\
\Delta y_{Nt} &= \delta_N + \alpha_N u_{t-1} + v_{Nt}, \\
\end{align*}
\] (6.9)

where the error correction term from the scalar cointegrating equation is now defined in terms of the equilibrium error

\[
u_t = y_{1t} - \beta_0 - \beta_2 y_{2t} - \beta_3 y_{3t} - \cdots - \beta_N y_{Nt}.
\]

(6.10)

A property of (6.10) is that the long-run parameter on \( y_{1t} \) is set at \( \beta_1 = 1 \), which is a normalisation that specifies \( y_{1t} \) as the dependent variable in this cointegrating equation. Such a normalization is usually specified because it has an explicit economic or financial interpretation. However, as all variables in a VECM are endogenous and jointly determined any normalisation that makes one of the \( N \) variables the dependent variable in the cointegrating equation is permissible, provided that variable does enter the cointegrating relation. So, for example if \( \beta_2 \neq 0 \), setting \( \beta_2 = 1 \) by rescaling the equation would make \( y_{2t} \) the dependent variable of the cointegrating equation. The scalar used in this rescaling is simply \( 1/\beta_1 \) and it is absorbed in each of the loading factors \( \alpha_i \) in the \( N \) equations. 

\footnote{This invariance of the cointegrating equation specification has implications for the finite sample properties of maximum likelihood estimators of the cointegrating parameters that are analogous to those in simultaneous equations models. In particular, these estimates typically have heavy tailed behavior in finite samples which increases the probability of outliers, as shown in Phillips (1994).}

**Additional \( I(0) \) Variables**

The VECM is so far specified in terms of \( I(1) \) variables. Stationary variables can also be included to contribute to the transient dynamics of the system. In the case of the present value model these additional determinants could be business cycle factors or external influences from world equity markets. As these additional variables are \( I(0) \) by definition the natural place for them to enter the model is in the error correction equations as these equations govern the short term movements in the nonstationary variables. To include a stationary variable \( x_t \) with an additional lag \( x_{t-1} \) into the bivariate VECM in (6.8), the model may be rewritten as

\[
\begin{align*}
\Delta y_{1t} &= \delta_1 + \alpha_1 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1}) + \psi_{11} x_t + \psi_{12} x_{t-1} + v_{1t}, \\
\Delta y_{2t} &= \delta_2 + \alpha_2 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1}) + \psi_{21} x_t + \psi_{22} x_{t-1} + v_{2t},
\end{align*}
\] (6.11)

where \( \psi_{ij} \) is the parameter on \( x_{t-j} \) in the \( i^{th} \) equation of the VECM. In this instance, the \( x_t \) variable is specified to enter all equations. A more restrictive choice is to allow \( x_t \) to enter only some of the VECM equations by setting the appropriate \( \psi_{ij} \) parameters to zero.
Short-run Dynamics

To allow the bivariate VECM in (6.8) to have additional short-run dynamics the model may be respecified as

\[
\Delta y_{1t} = \delta_1 + \alpha_1 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1}) + \sum_{i=1}^{k-1} \gamma_{11i} \Delta y_{1t-i} + \sum_{i=1}^{k-1} \gamma_{12i} \Delta y_{2t-i} + v_{1t},
\]

\[
\Delta y_{2t} = \delta_2 + \alpha_2 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1}) + \sum_{i=1}^{k-1} \gamma_{21i} \Delta y_{1t-i} + \sum_{i=1}^{k-1} \gamma_{22i} \Delta y_{2t-i} + v_{2t},
\]

(6.12)

where \( k \) controls the length of the lag structure in the transient dynamics. The strength of these additional dynamics in (6.12) are captured by the \( \gamma_{11i} \), \( \gamma_{12i} \), \( \gamma_{21i} \), and \( \gamma_{22i} \) parameters which measure the magnitude and direction of influence of the lagged dependent variables. The multivariate nature of a VECM means that all lagged dependent variables appear in all equations. Moreover, as these lags correspond to the changes in the endogenous variables \( \Delta y_{it} \), they affect the short-run dynamic path as the system adjusts towards its long-run equilibrium path. The short-run movements could be monotonic as is the case with the adjustment path from point B to point D in Figure 6.3, or they could now be more involved paths that embody cycles or even overshooting behaviour.

Time Trends

A further generalisation of the VECM in (6.12) is to include a time trend in the cointegrating equation as in the following specification

\[
\Delta y_{1t} = \delta_1 + \alpha_1 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1} - \phi t) + \sum_{i=1}^{k-1} \gamma_{11i} \Delta y_{1t-i} + \sum_{i=1}^{k-1} \gamma_{12i} \Delta y_{2t-i} + v_{1t},
\]

\[
\Delta y_{2t} = \delta_2 + \alpha_2 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1} - \phi t) + \sum_{i=1}^{k-1} \gamma_{21i} \Delta y_{1t-i} + \sum_{i=1}^{k-1} \gamma_{22i} \Delta y_{2t-i} + v_{2t},
\]

(6.13)

with the strength of the trend in the cointegrating equation controlled by the parameter \( \phi \). Imposing the restriction \( \phi = 0 \) eliminates the trend and reduces the VECM in (6.13) to (6.8). As a further extension a time trend could also be included in the error correction equations which, because the equation is formulated in differences, would result in the level variables \( y_{1t} \) and \( y_{2t} \) having quadratic deterministic time trends.
6.2. VECTOR ERROR CORRECTION MODELS

General Specification

The extensions discussed above are conveniently combined by expressing the VECM in matrix notation. Define the following \((N \times 1)\) vectors

\[
\begin{align*}
y_t &= \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{bmatrix}, \\
\delta &= \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_N \end{bmatrix}, \\
\alpha &= \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix}, \\
\beta &= \begin{bmatrix} 1 \\ -\beta_2 \\ \vdots \\ -\beta_N \end{bmatrix}, \\
v_t &= \begin{bmatrix} v_{1t} \\ v_{2t} \\ \vdots \\ v_{Nt} \end{bmatrix},
\end{align*}
\]

so that the VECM in (6.9) and (6.10) can be written as

\[
\Delta y_t = \delta + \alpha (\beta' y_{t-1} - \beta_0) + v_t, \quad (6.14)
\]

where \(\beta_0\) is a scalar. To capture the proposed extensions to the VECM, a general specification is given by

\[
\Delta y_t = \delta + \alpha (\beta' y_{t-1} - \beta_0 - \phi_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta y_{t-i} + \Psi x_t + v_t, \quad (6.15)
\]

where \(\phi\) is a scalar, \(\Gamma_i\) is a \((N \times N)\) matrix of parameters at lag \(i\), and \(\Psi\) is a \((N \times K)\) matrix of parameters associated with the \(K\) stationary variables contained in \(x_t\).

6.2.2 Relationship with VARs

An important property of a VECM is that it represents a restricted form of a VAR system. The restrictions arise from the property that the variables within the system are connected by the same long-run cointegrating equation. To highlight the nature of these restrictions consider the following bivariate VECM

\[
\begin{align*}
y_{1t} - y_{1t-1} &= \alpha_1 (y_{1t-1} - \beta_2 y_{2t-1}) + v_{1t}, \\
y_{2t} - y_{2t-1} &= \alpha_2 (y_{1t-1} - \beta_2 y_{2t-1}) + v_{2t}, \quad (6.16)
\end{align*}
\]

in which there is one cointegrating equation and no lagged difference terms on the right hand side. There are three parameters to be estimated, namely, the cointegrating parameter \(\beta_2\), and the two error correction parameters \(\alpha_1\) and \(\alpha_2\). Now re-express each equation in terms of the levels of the variables as

\[
\begin{align*}
y_{1t} &= (1 + \alpha_1) y_{1t-1} - \alpha_1 \beta_2 y_{2t-1} + v_{1t}, \\
y_{2t} &= \alpha_2 y_{1t-1} + (1 - \alpha_2 \beta_2) y_{2t-1} + v_{2t}, \quad (6.17)
\end{align*}
\]

or

\[
\begin{align*}
y_{1t} &= \phi_{11} y_{1t-1} + \phi_{12} y_{2t-1} + v_{1t}, \\
y_{2t} &= \phi_{21} y_{1t-1} + \phi_{22} y_{2t-1} + v_{2t}. \quad (6.18)
\end{align*}
\]

The parameters in (6.18) are related to those in (6.17) by the restrictions

\[
\phi_{11} = 1 + \alpha_1, \quad \phi_{12} = -\alpha_1 \beta_2, \quad \phi_{21} = \alpha_2, \quad \phi_{22} = 1 - \alpha_2 \beta_2. \quad (6.19)
\]
Equation (6.18) represents a VAR in the levels of the variables as discussed in Chapter 4. Although (6.17) and (6.18) look very similar (they both represent models containing the levels of the variables and the same number of lags) careful comparison of the VECM in (6.17) and the VAR in (6.18) reveals three details worth emphasising.

(i) The model in (6.17) is a restricted VAR containing 3 parameters whereas (6.18) is an unrestricted VAR containing 4 parameters. This difference in the number of unknown parameters is encapsulated by the set of restrictions in (6.19) which arise from the $y_{it}$ variables jointly having the same long-run equation.

(ii) If $y_{1t}$ and $y_{2t}$ are cointegrated then the VECM system specified in levels in (6.17) is embedded within the VAR model (6.18), suggesting that a VAR in levels is an appropriate general specification because it allows for both the VECM system by way of specialisation of the system as well as cases where there is no cointegration. When there is cointegration, the long-run and short run relationships manifest through the parameters on the lagged variables of the VAR in (6.18) that satisfy the restrictions (6.19). But when there is no cointegration between $y_{1t}$ and $y_{2t}$, the adjustment parameters $\alpha_1 = \alpha_2 = 0$ and there are no equilibrating forces present to return the system towards a long-run equilibrium. Imposing the restrictions $\alpha_1 = \alpha_2 = 0$ on (6.16) or (6.17) reduces the model to a simple VAR in first differences, that is a VAR model in which each equation has a unit root and there are no cointegrating links between the $I(1)$ variables. In such cases there are a full set of unit roots in the system and short-run dynamics can be modelled by examining the corresponding equation specified in differences.

(iii) The VAR in (6.18) contains one lag, whereas the VECM specification in (6.16) contains no additional lags of the dependent variable other than those embodied in the error correction component. This difference in specification of the lag structure extends to VARs with $k$ lagged dependent variables, with the corresponding VECM containing only $k - 1$ additional lagged dependent variables, all of which are specified in terms of differences. Thus, the number of additional lags in the VECM specification (6.15) is $k - 1$. Since those additional lags involve differences $\Delta y_{1-\ell}$, the corresponding unrestricted levels VAR specification has $k$ lags. This connection between the lag structures of the VAR and the VECM specifications is important in practical work. It means that the information criteria used in Chapter 4 to determine the optimal lag structure of a VAR are also valid for determining the optimal lag length of a VECM: if $k$ is the optimal lag structure in the levels VAR determined by information criteria, then the corresponding number of additional lags in differences in the VECM model is $k - 1$.

The possible relationships between VECM and VAR models may be sum-
marised using a bivariate VAR(1) model expressed as
\[ \Delta y_t = \Phi y_{t-1} + v_t. \] (6.20)
or in the equivalent form
\[ y_t = (I_2 + \Phi) y_{t-1} + v_t, \] (6.21)
in which \( I_2 \) is a 2 \times 2 identity matrix. Three versions of this model are possible, each of which involves a different rank of the matrix \( \Phi \), which is denoted by \( r \).

**Full rank case (\( r = 2 \))**
\[
\begin{bmatrix}
\Delta y_{1t} \\
\Delta y_{2t}
\end{bmatrix} =
\begin{bmatrix}
\phi_{11} & \phi_{12} \\
\phi_{21} & \phi_{22}
\end{bmatrix}
\begin{bmatrix}
y_{1t-1} \\
y_{2t-1}
\end{bmatrix} +
\begin{bmatrix}
v_{1t} \\
v_{2t}
\end{bmatrix}
\]

If \( \Phi \) is of full rank, \( r = 2 \), then all the elements of \( \Phi \) may be freely estimated without restriction. In this instance, the correct specification of the model is an unrestricted VAR in levels. The system is stationary and all variables are \( I(0) \) when the matrix \( (I_2 + \Phi) \) is stable.

**Reduced rank case (\( r = 1 \))**
\[
\begin{bmatrix}
\Delta y_{1t} \\
\Delta y_{2t}
\end{bmatrix} =
\begin{bmatrix}
\alpha_1 & \alpha_2 \\
\beta_1 & -\beta_2
\end{bmatrix}
\begin{bmatrix}
y_{1t-1} \\
y_{2t-1}
\end{bmatrix} +
\begin{bmatrix}
v_{1t} \\
v_{2t}
\end{bmatrix}
\]

In this case there is a cointegrating linear combination \( y_{1t-1} - \beta_2 y_{2t-1} = \beta'y_{t-1} \) between the two variables in levels and the matrix \( \Phi \) has the outer product form \( \Phi = \alpha\beta' \) where the vector \( \alpha' = (\alpha_1, \alpha_2) \) and the vector \( \beta' = (\beta_1, \beta_2) \). Examination of the matrix \( \Phi \) reveals that the first row is simply a multiple of the second row, so that its rank \( r = 1 \). Looked at another way, constraining \( \Phi \) to be the outer product of two vectors implies that the matrix has at most the rank of its constituent parts, which both have rank \( r = 1 \). The model may be specified as a restricted VAR expressed in levels or as a VECM expressed in differences and levels. The concept of reduced rank is used in the next two sections in discussing the estimation and testing of cointegrated systems of equations.

**Zero rank case (\( r = 0 \))**
\[
\begin{bmatrix}
\Delta y_{1t} \\
\Delta y_{2t}
\end{bmatrix} =
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
y_{1t-1} \\
y_{2t-1}
\end{bmatrix} +
\begin{bmatrix}
v_{1t} \\
v_{2t}
\end{bmatrix}
\]
The most restricted model is the VAR in first differences given by the third model. This model has zero rank as the coefficient matrix \( \Phi \) is merely a matrix of zeros, the outcome of four coefficient restrictions. In this case the model is a VAR in first differences, each variable is \( I(1) \), and the system is sometimes called a full rank \( I(1) \) VAR model.

---

3Appendix A provides a short introduction to matrices including discussion of the identity matrix and the rank of a matrix.
CHAPTER 6. COINTEGRATION

6.3 Estimation

Two methods for estimating the unknown parameters of the VECM introduced in Section 6.2 are now discussed. These are the fully modified (FM-OLS) regression-based estimator proposed by Phillips and Hansen (Phillips and Hansen, 1990) and the Johansen reduced rank regression estimator (Johansen, 1988, 1991, 1995).

These estimators both possess two important properties.

(i) They are consistent for the true population parameters at the rate $T$, which is considerably faster than the $\sqrt{T}$ rate at which coefficients are estimated in stationary time series models (see, for example, Appendix B). This faster rate of convergence is sometimes referred to as super-consistency. It arises because the variables $\{y_{1t}, y_{2t}, \cdots, y_{Nt}\}$ are non-stationary $I(1)$ series, whose stochastically trending nature facilitates estimation of the cointegrating relationships.

(ii) The asymptotic distributions of these two estimators of the cointegrating parameters can be used to conduct hypothesis tests in much the same way as for models based on stationary variables. This important feature is not true for all cointegrating regression estimators.

6.3.1 The Fully Modified Estimator

The FM-OLS estimator of Phillips and Hansen (Phillips and Hansen, 1990) is a regression based estimator that circumvents the problems associated with the simple least squares regression estimation (Engle and Granger, 1987) by appropriately modifying the ordinary least squares estimator of the cointegrating equation. To highlight the features of the FM estimator, the present value model is specified more completely in terms of the following system:

\[
\begin{align*}
    p_t &= \beta_0 + \beta_d d_t + u_t, \\
    \Delta d_t &= v_t. 
\end{align*}
\]  

\[\text{(6.22)}\]

\[\text{---}\]

\[\text{4}\] Other methods for obtaining estimates of long-run cointegrating parameters have been proposed by Engle and Yoo, (1991), Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993). Most recently, methods that rely on deterministic trend instrumental variable regressions have been introduced by Phillips (2014b) and Hwang and Sun (2017). These latter methods have the advantage of simplicity and have good performance in both estimation and inference.

\[\text{5}\] The least squares estimator of the cointegrating parameters which were were used in the original study by Engle and Granger (1987) and for estimating the parameters of the present value model in equation (6.7) has the super-consistency property but the standard errors of the estimator do not allow conventional inference.

\[\text{6}\] FM-OLS may also be used in a system of equations and in a VAR context, where it is called FM-VAR (Phillips, 1995)
6.3. ESTIMATION

The first equation is the cointegrating equation as given in (6.1), while the second equation makes explicit the unit root stochastic trend property of dividends. The disturbances $u_t$ and $v_t$ are $I(0)$: the former because of cointegration and the latter because dividends are assumed to be difference stationary.

Although estimation of the cointegrating equation in (6.22) by ordinary least squares delivers super-consistent parameter estimates, inferences based on these estimates are generally invalid for several reasons:

(i) $d_t$ is an endogenous regressor;
(ii) $u_t$ and $v_t$ are correlated;
(iii) $u_t$ and $v_t$ are autocorrelated;

The FM estimator modifies least squares by taking into account these general features of the generating mechanism of the data. The endogeneity is addressed by making a correction for endogeneity and the autocorrelation is addressed by means of a serial correlation correction. In both cases long-run covariances are used to implement the corrections and these involve multivariate extensions of the long-run variance estimator introduced in Chapter 5. Long-run covariances are required in these corrections because the equation of interest in (6.22) is a long-run relationship.

The long-run matrices that are needed are defined in terms of the following two-sided and one-sided sums of autocovariances.

$$
\Omega = \begin{bmatrix}
\omega_{11} & \omega_{12} \\
\omega_{21} & \omega_{22}
\end{bmatrix} = \sum_{j=-\infty}^{\infty} \Gamma_j \quad \text{[2-sided]}
$$

$$
\Lambda = \begin{bmatrix}
\lambda_{11} & \lambda_{12} \\
\lambda_{21} & \lambda_{22}
\end{bmatrix} = \sum_{j=0}^{\infty} \Gamma_j \quad \text{[1-sided]}
$$

where $\Gamma_j = E(w_tw_{t-j}')$ with $w_t = (u_t, v_t)'$.

The required estimates of the long-run quantities are obtained by first estimating (6.22) by least squares. Letting $\hat{u}_t$ and $\hat{v}_t = \Delta d_t = v_t$ be the least squares residuals from estimating (6.1), the two-sided and one-sided long-run covariance matrices can be estimated by the following weighted sums of sample autocovariances

$$
\hat{\Omega} = \begin{bmatrix}
\hat{\omega}_{11} & \hat{\omega}_{12} \\
\hat{\omega}_{21} & \hat{\omega}_{22}
\end{bmatrix} = \sum_{j=-m}^{m} \left(1 - \frac{j}{m+1}\right) \hat{\Gamma}_j \quad \text{[2-sided]}
$$

$$
\hat{\Lambda} = \begin{bmatrix}
\hat{\lambda}_{11} & \hat{\lambda}_{12} \\
\hat{\lambda}_{21} & \hat{\lambda}_{22}
\end{bmatrix} = \sum_{j=0}^{m} \left(1 - \frac{j}{m+1}\right) \hat{\Gamma}_j \quad \text{[1-sided]}
$$

where $m$ is the lag length and $\hat{\Gamma}_j$ is the autocovariance matrix at lag $j$ given by

$$
\hat{\Gamma}_j = \frac{1}{T} \sum_{t=1+j}^{T} \begin{bmatrix}
\hat{u}_t\hat{u}_{t-j} & \hat{u}_t\hat{v}_{t-j} \\
\hat{v}_t\hat{u}_{t-j} & \hat{v}_t\hat{v}_{t-j}
\end{bmatrix}.
$$
The diagonal elements of \( \hat{\Gamma}_j \) are the sample autocovariances of \( \hat{u}_t \) and \( \hat{v}_t \), and the off-diagonal elements contain their sample cross-covariances. The weighting scheme used in (6.22) employs the triangular weights \( (1 - j)/(m + 1) \) and was introduced by Bartlett (1950) and used by Newey and West (1987) in the development of robust methods for constructing standard errors in econometrics.

The FM estimator of the cointegrating equation in (6.22) corrects for endogeneity and serial dependence and has the following explicit form

\[
\hat{\beta}_{FM} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_d \end{bmatrix} = \left( \frac{T}{\sum_{t=1}^{T} d_t} \right)^{-1} \left( \frac{T}{\sum_{t=1}^{T} p_t^+} \right) \left( \frac{T}{\sum_{t=1}^{T} (d_t p_t^+ - \hat{\omega})} \right), \tag{6.26}
\]

where \( p_t^+ = p_t - \hat{\rho} \Delta d_t \) and \( \hat{\omega} = \hat{\omega}_2 \hat{\omega}_{22}^{-1} \). There are two correction terms in (6.26) that modify the usual ordinary least squares formula.

(i) The estimator \( \hat{\rho} = \hat{\omega}_{12} \hat{\omega}_{22}^{-1} \) is an estimator of the long-run regression coefficient \( \omega_{12} \omega_{22}^{-1} \) of \( u_t \) on \( v_t \). The correction term \( p_t^+ = p_t - \hat{\rho} \Delta d_t \) then adjusts the dependent variable \( p_t \) for its long-run joint dependence on \( d_t \). This is the endogeneity correction.

(ii) The quantity \( \hat{\omega} = \hat{\omega}_{12} \hat{\omega}_{22}^{-1} \) provides an adjustment that compensates for the contemporaneous and serial dependence of the price equation errors \( u_t \) with the errors \( \Delta d_t = v_t \) in the dividend equation. This adjustment leads to the modification of sample moment \( \sum_{t=1}^{T} (d_t p_t^+ - \hat{\omega}) \) that appears in the final member of the right side of (6.26).

Combining the adjustments (i) and (ii) leads to the fully modified version of least squares regression. To help understand the properties of the FM estimator consider the special case where the equation errors \( u_t \) and \( v_t \) are not cross-correlated at lags and leads. The estimator of the covariance matrix in (6.25) is computed just using the sample contemporaneous covariance matrix between the \( \hat{u}_t \) and \( \hat{v}_t \) since \( \Gamma_j = 0 \) for all \( j \neq 0 \). For this special case the estimators of the long-run covariance matrices in (6.24) are equal to each other

\[
\begin{bmatrix} \hat{\omega}_{11} \\ \hat{\omega}_{21} \\ \hat{\omega}_{22} \end{bmatrix} = \begin{bmatrix} \hat{\lambda}_{11} \\ \hat{\lambda}_{21} \\ \hat{\lambda}_{22} \end{bmatrix} = \frac{1}{T} \sum_{t=1}^{T} \begin{bmatrix} \hat{u}_t^2 \\ \hat{v}_t \hat{u}_t \\ \hat{v}_t^2 \end{bmatrix},
\]

which results in the correction term \( \hat{\omega} \) in (6.26) disappearing since

\[
\hat{\omega} = \hat{\lambda}_{12} \hat{\lambda}_{22}^{-1} = \hat{\omega}_{12} - \hat{\rho} \hat{\omega}_{22} = \hat{\omega}_{12} - \hat{\omega}_{12} \hat{\omega}_{22}^{-1} \hat{\omega}_{22} = 0.
\]

The FM estimator is now simply obtained by regressing the modified dependent variable \( p_t^+ = p_t - \hat{\rho} \Delta d_t \) on a constant and \( d_t \), with the remaining correction term \( \hat{\rho} \) computed from a simple least squares regression of \( \hat{u}_t \) and \( \hat{v}_t \).
A further special case is where there is no contemporaneous or serial correlation between \( u_t \) and \( v_t \) at all in (6.22) so the appropriate estimator for the remaining correction term is just \( \hat{\rho} = 0 \). In this case, where there is no endogeneity and no serial correlation and this information is used, there is equivalence of the FM-OLS and ordinary least squares estimates. In general, of course, endogeneity and serial correlation are to be expected in the data and in the absence of any explicit information to the contrary, FM-OLS is constructed using both corrections.

The FM-OLS estimator is implemented by the following steps.

**Step 1:** Estimate equation (6.22) by ordinary least squares to obtain \( \hat{\beta}_0 \) and \( \hat{\beta}_d \) and the residuals \( \hat{u}_t \), as well as \( \hat{v}_t = v_t = \Delta d_t \).

**Step 2:** Compute the long-run covariance matrices in (6.24) using the matrices in (6.25) with the lag length \( m \) selected according to some fixed value or data-determined according to some optimal selection criterion.

**Step 3:** Use (6.26) to compute the FM parameter estimates.

Applying these steps to the United States data on equity prices and dividends presented in Figure 6.1 gives the FM estimates of the present value cointegrating equation with standard errors in parentheses

\[
p_t = 3.127 + 1.195 d_t + \hat{u}_t. \tag{6.27}
\]

The long-run covariance matrices in (6.24) are computed using a lag length of \( m = 8 \). The FM-OLS parameter estimates are qualitatively very similar to the OLS parameter estimates presented in equation (6.2) which reflects the fact that both estimators generate super-consistent parameter estimates. However, the standard errors associated with the FM-OLS estimates given in (6.27) are robust to endogeneity and serial dependence by virtue of their construction.

### 6.3.2 The Johansen Reduced Rank Regression Estimator

The reduced rank regression estimator (Johansen, 1988, 1991, 1995) yields estimates for all of the parameters in an explicit parametric form of the VECM in (6.15), including both the long-run cointegrating parameters and the short-run error-correction parameters. This is in contrast to the semi-parametric regression-based FM estimator which yields estimates of only the long-run parameters because the transient dynamics are treated nonparametrically by FM-OLS.

To demonstrate the operation of the Johansen estimator, consider the following VECM of equity prices and dividends with the short-run dynamics cap-
tured by an additional lagged difference in the two variables

\[ p_t = \beta_0 + \beta_d d_t + u_t, \]

\[ \Delta p_t = \delta_1 + a_1(p_{t-1} - \beta_0 - \beta_d d_{t-1}) + \gamma_{11} \Delta p_{t-1} + \gamma_{12} \Delta d_{t-1} + v_{1t}, \]  

\[ \Delta d_t = \delta_2 + a_2(p_{t-1} - \beta_0 - \beta_d d_{t-1}) + \gamma_{21} \Delta p_{t-1} + \gamma_{22} \Delta d_{t-1} + v_{2t}, \] 

(6.28)

where \( v_t = (v_{1t}, v_{2t})' \) is a \((2 \times 1)\) vector of the VECM disturbances. As pointed out in Section 6.2 a VECM is a restricted VAR model in which the restrictions arise from a common long-run equation relation that leads to a reduced rank VAR structure. From inspection of the VECM in (6.28) the cointegrating link between prices and dividends yields a set of nonlinear cross-equation restrictions in which the cointegrating parameters \( \beta_0 \) and \( \beta_d \) appear in both equations.

Johansen develops a computationally efficient algorithm to estimate the parameters of the VECM using reduced rank regression under the assumption of Gaussian innovations.\(^7\) The Johansen estimator is implemented using the following steps.

**Step 1:** Construct the conditional residuals from 4 auxiliary regressions by regressing \( \{\Delta p_t, \Delta d_t, p_{t-1}, d_{t-1}\} \) on \( \{1, \Delta p_{t-1}, \Delta d_{t-1}\} \).

**Step 2:** Use the conditional residuals in the previous step to perform an eigen decomposition and compute \( \hat{\beta}_0 \) and \( \hat{\beta}_d \) using the eigenvector corresponding to the largest eigenvalue.\(^8\)

**Step 3:** Construct the cointegrating residuals \( \hat{u}_t = p_t - \hat{\beta}_0 - \hat{\beta}_d d_t \) and estimate the VECM in (6.28) by regressing \( \{\Delta p_t, \Delta d_t\} \) in turn, on the regressors \( \{1, \hat{u}_{t-1}, \Delta p_{t-1}, \Delta d_{t-1}\} \).

The Johansen parameter estimates are presented in Table 6.1 using the United States data on equity prices and dividends presented in Figure 6.1. The cointegrating parameter estimate of \( \beta_d \) is 1.1773, which is marginally smaller that the FM estimate of 1.195 given in equation (6.27). The signs of the error-correction parameters are \( \hat{\alpha}_1 = -0.0068 < 0 \) and \( \hat{\alpha}_2 = 0.0024 > 0 \). These signs are consistent with the system adjusting towards its long-run equilibrium by means of the price and dividend changes discussed earlier. The results are similar qualitatively to those obtained in (6.7) where the cointegrating residual \( \hat{u}_t \) is based on the OLS estimator.

The estimate of the intercept parameter \( \beta_0 \) does not have an associated standard error, whereas the estimates of the intercepts in the 2 error correction equations do have standard errors. The reason for this difference is that the

---

\(^7\)The Johansen estimator is based on the maximum likelihood principle, which is introduced in Chapter 10. The treatment here does not emphasise this aspect of the estimator.

\(^8\)The eigenvalue decomposition is also used to estimate factor models based on principal components in Chapter 12.
6.3. ESTIMATION

dimension of the system is $N = 2$ and there are therefore only 2 unique intercepts corresponding to the levels of the 2 variables $p_t$ and $d_t$. The reduced rank regression estimator apportions these values among the 3 intercepts that appear in the model $(\beta_0, \gamma_1$ and $\gamma_2)$ thereby delivering estimates of the 3 intercepts. Only 2 standard errors are computed and provided because there are only two random intercepts from which the 3 intercepts in the VECM system are obtained.

Table 6.1

The parameter estimates of the price-dividend VECM in (6.28) with standard errors in parentheses. The sample period is January 1871 to September 2016. The estimated cointegrating equation is

$$p_t = 3.4034 + 1.1773 d_t + \hat{u}_t.$$  

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\Delta p_t$</th>
<th>$\Delta d_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0004</td>
<td>0.0010</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0002)</td>
</tr>
<tr>
<td>$\hat{u}_{t-1}$</td>
<td>-0.0068</td>
<td>0.0024</td>
</tr>
<tr>
<td></td>
<td>(0.0032)</td>
<td>(0.0004)</td>
</tr>
<tr>
<td>$\Delta p_{t-1}$</td>
<td>0.2901</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>(0.0231)</td>
<td>(0.0030)</td>
</tr>
<tr>
<td>$\Delta d_{t-1}$</td>
<td>0.1366</td>
<td>0.8792</td>
</tr>
<tr>
<td></td>
<td>(0.0841)</td>
<td>(0.0110)</td>
</tr>
</tbody>
</table>

From the theory underlying the present value model, equation (2.7) in Chapter 2 shows that the intercept $\beta_0$ in the cointegration equation of the present value model can be used to generate an estimate of the implied discount factor. This estimate is

$$\exp(-\hat{\beta}_0) = \exp(-3.4034) = 0.0333,$$

resulting in a discount factor of 3.33% which is just slightly below the estimate of 5% obtained from Figure 2.5 in Chapter 2.

The parameter estimates of the present value model presented in Table 6.1 are based on specifying the VECM in (6.15) with a single lag. This choice is arbitrary and is made here to simplify the implementation and the presentation of the results for the present value model. Formal data-based approaches to select the lag structure may be used instead. A simple method is to estimate the unrestricted VAR form of a VECM that is discussed in Section 6.2 and use information criteria to determine an appropriate lag length $k$ of the VAR. The corresponding lag length for the VECM, where the lags are formulated in differences, is $k - 1$. A preferred approach is to use information criteria to select both the lag length in the reduced rank regression and the
cointegrating rank simultaneously by information criteria. This approach was suggested by Phillips (1996) and yields consistent estimates of both lag length and cointegrating rank (Chao and Phillips, 1999). More recent work (Cheng and Phillips, 2009; Cheng and Phillips, 2012) has shown that information criteria may be used to consistently estimate cointegrating rank without requiring explicit specification of the transient dynamics in the fitted model.

Determination of an appropriate lag structure is very important in implementing reduced rank regression in practice. The parameter estimates of the cointegrating equation and the choice of cointegrating rank (or the number of cointegrating relations) can both be extremely sensitive to the choice of \( k \). This fragility is particularly the case in large dimensional systems where a large loss in degrees of freedom occurs as \( k \) increases, which in turn leads to imprecise parameter estimates. In such cases it is desirable to employ data-determined methods for joint determination of lag length and cointegrating rank. Another strategy is to employ semi-parametric methods such as FM-OLS regression, which has been found to be more robust in estimating the cointegrating equation than reduced rank regression estimation in which lag structure is determined parametrically. The sensitivities of reduced rank regression to choice of lag length in the present value VECM are explored in the exercises.

### 6.4 Cointegration Testing

Both the FM-OLS and Johansen estimators discussed in Section 6.3 proceed under the assumption that a cointegrating long-run relationship between the \( I(1) \) variables exists. For the bivariate model relating equity prices and dividends the long-run relationship originates in the present value finance model of equity prices and the presumption made in estimation is that this relationship is satisfied empirically.\(^9\)

In this section two methods are proposed to test for the presence of cointegration. The first method is easy to conduct and applies the univariate unit root tests of Chapter 5 to the residuals obtained from estimating the cointegrating equation. The second method relies on reduced-rank regression and may be interpreted as a multivariate extension of the univariate unit root test.

\(^9\)If no such long-run relationship exists, then empirical regressions of \( I(1) \) variables which are not cointegrated produce spurious results in the sense that the variables may appear to be related because of the existence of their individual underlying trends even though these trends are not common to both variables. This problem is known as the spurious regression problem. In cases where there is no underlying long-run relationship, an unrestricted VAR in first differences (instead of a VECM formulation) may still be employed and can deliver valid inference about the short-run dynamics. Valid inference may also be made in VAR regressions formulated in levels provided appropriate methods are used, such as FM-OLS regression (Phillips, 1995) or lag augmented VAR regression (Toda and Yamamoto, 1995).
6.4. COINTEGRATION TESTING

6.4.1 Residual-based Tests

To illustrate the residual-based test of cointegration consider the present value model

\[ p_t = \beta_0 + \beta_d d_t + u_t, \]  
(6.29)

where the log equity price \( p_t \), and log dividend \( d_t \) are \( I(1) \) and \( u_t \) is the disturbance term which will be \( I(0) \) when \( p_t \) and \( d_t \) are cointegrated. A natural way to test for cointegration is a two-step procedure consisting of estimating the cointegrating equation in (6.29) by ordinary least squares in the first step and testing the residuals for stationarity in the second step. As the unit root test treats the null hypothesis as nonstationary, the null hypothesis corresponds to no cointegration whereas the alternative hypothesis of stationarity corresponds to cointegration:

\[ H_0 : \text{No Cointegration} \quad [u_t \text{ is nonstationary}] \]
\[ H_1 : \text{Cointegration} \quad [u_t \text{ is stationary}] \]  
(6.30)

The distribution of the residual-based cointegration test is non-standard, as might be expected because the procedure tests for a unit root null hypothesis (Phillips and Ouliaris, 1990). In a similar way to unit root tests, residual-based cointegration tests have critical values that depend on the sample size and the number of deterministic terms in the regression. But unlike standard unit root tests, residual based tests have asymptotic distributions and critical values that also depend on the number of \( I(1) \) regressors in the cointegrating equation.\(^\text{10}\) Tables are provided by MacKinnon (1991) which gives response surface estimates of the critical values that are now used in many computer packages.

Using the monthly data on equity prices and dividends for the United States from January 1871 to September 2016, the residuals from estimating the cointegrating equation in (6.29) by ordinary least squares are plotted in Figure 6.4. The series has mean zero and there is no trend apparent, giving the appearance of stationarity.

Formal tests of the stationarity of the residuals in Figure 6.4 are presented in Table 6.2. Two cointegration tests are provided based on the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) unit root tests with lag lengths running from 0 to 4 lags. The ADF tests based on lags 1 to 4 all yield \( p \) values less than 0.05 thereby providing strong evidence for rejecting the null of no cointegration in favour of the alternative hypothesis of cointegration. Similar qualitative results are obtained for the PP test for all lags. These results in general confirm the intuition provided by Figure 6.4 thereby providing strong support for the present value model as a valid long-run relationship between equity prices and dividends.

\(^{10}\)No degrees of freedom adjustment is necessary if the cointegrating parameters are known, as is the case for the present value model if the long-run restriction of \( \beta_d = 1 \) is imposed on (6.29). In this case the critical values used to test for a unit root in a variable are still appropriate.
CHAPTER 6. COINTEGRATION

Figure 6.4: Plot of the residuals from the first stage of the Engle-Granger two stage procedure applied to the dividend model. Data are monthly observations from February 1871 to September 2016 on United States equity prices and dividends.

Table 6.2
Residual-based tests of cointegration between United States equity prices and dividends using equation (6.29). The test regression has no constant term with lags running from 0 to 4.

<table>
<thead>
<tr>
<th>Lags</th>
<th>Augmented Dickey-Fuller</th>
<th>Phillips-Perron</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>−3.028</td>
<td>−3.028</td>
</tr>
<tr>
<td>1</td>
<td>−4.287</td>
<td>−3.493</td>
</tr>
<tr>
<td>2</td>
<td>−4.029</td>
<td>−3.680</td>
</tr>
<tr>
<td>3</td>
<td>−4.037</td>
<td>−3.770</td>
</tr>
<tr>
<td>4</td>
<td>−4.317</td>
<td>−3.849</td>
</tr>
</tbody>
</table>

(i) Critical values are, respectively: 1% −2.580; 5% −1.950; and 10% −1.620.

(ii) In the ADF test, lags(·) specifies the number of lagged differences of the dependent variable included in the regression. In the Phillips-Perron test, lags(·) specifies the number of Newey-West lags employed in the estimates of the nonparametric components of the test statistic.
6.4. COINTEGRATION TESTING

6.4.2 Johansen Reduced Rank Regression Tests

The Johansen reduced rank regression test of cointegration is based on the VECM, which for the present value model is given by the two error correction equations for $p_t$ and $d_t$ as

\[
\Delta p_t = \delta_1 + \alpha_1 (p_{t-1} - \beta_0 + \beta_d d_{t-1}) + \gamma_{11} \Delta p_{t-1} + \gamma_{12} \Delta d_{t-1} + \nu_{1t},
\]

\[
\Delta d_t = \delta_2 + \alpha_2 (p_{t-1} - \beta_0 + \beta_d d_{t-1}) + \gamma_{21} \Delta p_{t-1} + \gamma_{22} \Delta d_{t-1} + \nu_{2t}.
\] (6.31)

Unlike the residual-based tests of cointegration which involve testing a single set of hypotheses as in (6.30), the Johansen test is based on testing a sequence of different hypotheses. In the case of the present value model where there are $N = 2$ equations, there are two stages to the testing procedure with different null and alternative hypotheses in each case:

**Stage 1**
- $H_0$: No cointegration, all variables are $I(1)$
- $H_1$: 1 or more cointegrating equations

**Stage 2**
- $H_0$: 1 cointegrating equation
- $H_1$: All variables are stationary $I(0)$

The first stage is effectively equivalent to the null and alternative hypotheses used in (6.30). The null hypothesis is no cointegration which occurs when $p_t$ and $d_t$ are $I(1)$ and the error correction term $u_t = p_t - \beta_0 + \beta_d d_t$ is also $I(1)$. The alternative hypothesis at the first stage is that cointegration exists and $u_t$ is $I(0)$. If the null hypothesis in the first stage is not rejected testing is complete with the conclusion that there is no cointegration.

If the first stage null hypothesis is rejected in favour of the alternative hypothesis in (6.32) the second stage hypotheses are tested. Here the null hypothesis is that there is one cointegrating equation and so $u_t$ is $I(0)$. Interestingly, the alternative hypothesis for this second stage is not that there are two cointegrating equations, but that $p_t$ and $d_t$ are actually stationary or $I(0)$.

A simple way of interpreting this sequence of tests is to view them as a test for determining the number of unit roots in the system or the number of unit roots in the VAR model expressed in levels. The null hypothesis in Stage 1 is that there is a full set of unit roots. The alternative hypothesis in Stage 1 is that there are at most $N - 1$ unit roots (or at least one cointegrating relation).

The null hypothesis in Stage 2 is that there are $N - 1$ unit roots and one cointegrating relation. The alternative hypothesis in Stage 2 is that there are $N - 2$ unit roots and, since $N = 2$ in this example, the system is stationary. Thus, the reduced rank regression test sequence is really just a sequence of tests concerning the number of unit roots in the system, thereby making these tests just a multivariate form of unit root test.

To test the sequence of hypotheses in (6.32), the first, most restrictive case involves estimating the VECM in (6.31) under the null hypothesis of no cointegration ($r = 0$) in Stage 1 of (6.32) by restricting the error correction parameters to be zero $\alpha_1 = \alpha_2 = 0$. The VECM in (6.31) then reduces to a VAR with
CHAPTER 6. COINTEGRATION

$k - 1$ lags with all of the variables in this case expressed in first differences. This is a VAR with a full set of unit roots. The second set of restrictions corresponds to the alternative hypothesis in the first stage in (6.32), equivalent to the null hypothesis in the second stage, by estimating the VECM in (6.32) subject to the cross-equation restrictions arising from cointegration ($r = 1$). The third set corresponds to the alternative hypothesis in the second stage where all variables are stationary ($r = 2$). In this scenario the VECM becomes a VAR with $k$ lags with all of the variables in levels. Since this approach to testing for cointegration is based on the rank of the system it is referred to as a reduced rank regression test or a Johansen test after (1988).

A computationally efficient way of performing the sequence of tests is to express the statistic in terms of the estimated eigenvalues obtained from applying the reduced rank regression estimator to the VECM in (6.15) (Johansen, 1988,1991,1995). For an $N$-dimensional system the form of the statistic is

$$\text{TRACE} = -(T - k) \sum_{i=r+1}^{N} \log(1 - \hat{\lambda}_i),$$

(6.33)

where $\hat{\lambda}_i, i = 1, 2, \cdots, N$ are the eigenvalues obtained from estimating the VECM using the reduced rank regression estimator, ordered from highest to lowest.\footnote{This test is, in fact, a likelihood ratio (LR) test since the reduced rank estimator is a maximum likelihood estimator under Gaussian assumptions. This type of test is dealt with in Chapter 10. However, unlike the standard chi-square distribution of the usual LR statistic, the trace test has a nonstandard distribution due to the nonstationarity of the variables being tested.} The subscript $r$ represents the rank of the system. Under the null hypothesis in Stage 1 of (6.32) the rank of the system is $r = 0$ with the statistic in (6.33) representing a joint test that all $N$ eigenvalues are zero. The null hypothesis in Stage 2 is that $r = 1$. The statistic in this case is a test that the smallest eigenvalue is zero. The test statistic is called the trace test because the trace of a matrix is determined by the number of non-zero eigenvalues. Large values of the trace statistic relative to the critical value result in rejection of the null hypothesis.

In the present value model $N = 2$ and therefore a sequence of two stages as in (6.32) is sufficient. For higher dimensional VECMs with $N > 2$, there may be more than a single cointegrating long-run equation connecting the nonstationary variables up to a possible maximum of $N - 1$ cointegrating equations. Examples of VECMs with multiple cointegrating equations are discussed in Sections 6.6 ($r = 3$) and 6.7 ($r = 5$).

The results of the Johansen test applied to the United States equity prices and dividends from January 1871 to September 2016 ($T = 1749$) are given in Table 6.3. The estimates of the eigenvalues are $\hat{\lambda}_1 = 0.0207$ and $\hat{\lambda}_2 = 0.0008$. The value of the test statistic for the first null hypothesis of no cointegration ($r = 0$) is computed as

$$\text{TRACE}(r = 0) = -1747 (\log(1 - 0.0207) + \log(1 - 0.0008)) = 37.9330.$$
The null is easily rejected at the 5% level with a $p$ value of 0.0000, providing evidence of at least one cointegrating equation between equity prices and dividends. The value of the test statistic for the next null hypothesis corresponding to one cointegrating equation ($r = 1$) is

\[ \text{TRACE}(r = 1) = -1747(\log(1 - 0.0008)) = 1.3910. \]

As the $p$ value is 0.2382 which is greater than 0.05, this null is not rejected at the 5% level. This sequence of tests provides strong support for the present value model because cointegration between equity prices and dividends is confirmed, thereby complementing the results obtained using the residual-based tests for cointegration.

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Eigenvalue</th>
<th>Statistic</th>
<th>5% CV</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r = 0$: No cointegration</td>
<td>0.0207</td>
<td>37.9330</td>
<td>15.41</td>
<td>0.000</td>
</tr>
<tr>
<td>$r = 1$: 1 cointegrating equation</td>
<td>0.0008</td>
<td>1.3910</td>
<td>3.76</td>
<td>0.238</td>
</tr>
</tbody>
</table>

6.4.3 Relationship with Unit Root Testing

As indicated in the previous discussion, the reduced-rank regression test for cointegration is a form of multivariate unit root test in which the variables are tested jointly for nonstationarity. If the null hypothesis in Stage 1 of (6.32) is not rejected then testing concludes with the result that all variables in the system are $I(1)$. This test may be compared with the approach in Chapter 5 where unit root tests are performed on each series at a time. In fact, the asymptotic distribution of the reduced-rank test is the analogue for a multivariate time series of the scalar unit root limit distribution for a single time series.

In the extreme case where all null hypotheses in (6.32) are rejected the final alternative hypothesis that all variables are $I(0)$ is accepted. If $N = 1$, this set of hypotheses reduces to the set of hypotheses underlying unit root tests discussed in Chapter 5, with the intermediate hypotheses of cointegration no longer relevant. A system that involves $N > 1$ variables allows for intermediate cases of cointegration where linear subsets of $I(1)$ variables can be constructed that are $I(0)$. If there are $r$ linearly independent subsets of this type, then there are still $N - r$ unit roots in the system and the reduced-rank regression test is a test for the presence of such unit roots.

The Johansen test provides a computationally efficient mechanism for computing unit root tests on several variables jointly with a single statistic. It also
circumvents situations where individual unit root tests may not provide a consensus view of the nonstationarity properties of the variables in the system. Cointegration testing may be interpreted as a step towards formulating a multivariate dynamic model of the VECM form. In doing so, the approach works to preserve the potential multivariate linkages between variables and thereby helps to avoid possible misspecification that can arise in making inference based on univariate specifications alone.

6.5 Parameter Testing

When there is cointegration in the system, the FM-OLS and reduced-rank regression estimators deliver super consistent parameter estimates of the cointegrating parameters associated with the $I(1)$ variables together with standard errors that can be used to construct asymptotically valid $t$ tests concerning these parameters. In contrast to the unit root tests discussed in Section 5 and the reduced rank regression tests for cointegration discussed above, both of which have non-standard asymptotic distributions, the distributions of these $t$ tests are, in fact, asymptotically normally distributed as $N(0,1)$ despite the component variables being nonstationary. This property arises because when it is correctly specified the VECM system is expressed in stationary variable form involving either first differences of $I(1)$ variables or long-run equilibrium errors such as $u_{t-1}$.

In testing the parameters of a VECM three broad sets of hypotheses can be explored. These hypotheses relate to the cointegrating parameters, the equilibrium adjustment parameters, and the transient dynamic parameters. The tests are illustrated for the present value cointegrating model using the reduced-rank regression parameter estimates in Table 6.1.

Cointegrating Parameters

The long-run parameter estimate on dividends in the cointegrating equation in (6.28) is $\hat{\beta}_d = 1.1773$, which is numerically close to the theoretical value predicted by the present value model of $\beta_d = 1$. Testing this restriction that $\beta_d = 1$ proceeds under the null hypothesis as usual and is based on the $t$ statistic

$$t(\beta_d = 1) = \frac{1.1773 - 1}{0.0303} = 5.8515.$$  

Despite the point estimate being numerically close to the theoretical value of $\beta_d = 1$ the large value of the $t$ statistic provides strong evidence that the null hypothesis is rejected at the 5% level.
6.5. PARAMETER TESTING

Error Correction Parameters and Weak Exogeneity

The relative strengths of the equilibrating mechanisms are controlled by the adjustment or error correction parameters \( \alpha_1 \) and \( \alpha_2 \) in (6.28). A test that \( p_t \) does not adjust to restore equilibrium is based on the null hypothesis that \( \alpha_1 = 0 \). This test is sometimes referred to as a test of weak exogeneity because under the null hypothesis movements in \( p_t \) are independent of the error correction term \( u_{t-1} = p_{t-1} - \beta_0 - \beta_d d_{t-1} \) and hence the long-run parameters \( \beta_0 \) and \( \beta_d \). The \( t \) statistic to test for weak exogeneity of \( p_t \) is

\[
t(\alpha_1 = 0) = \frac{-0.0068 - 0}{0.0032} = -2.1015.
\]

The \( p \) value is 0.0356 resulting in a rejection of the null hypothesis at the 5% level of significance, but not at the 1% level. A similar test of weak exogeneity is performed for dividends where the null hypothesis is \( \alpha_2 = 0 \). The \( t \) statistic is

\[
t(\alpha_2 = 0) = \frac{0.0024 - 0}{0.0004} = 5.7796,
\]

resulting in a \( p \) value of 0.000 and strong rejection of the null hypothesis. The outcome of these tests is that at the 5% level both \( p_t \) and \( d_t \) are important in adjusting to restore long-run equilibrium, whereas at the 1% level the primary role in the adjustment mechanism is taken by \( d_t \).

Dynamic Parameters and Strong Exogeneity

The short-run parameters \( \gamma_{ij} \) in (6.28) combined with the error correction parameters, \( \alpha_1 \) and \( \alpha_2 \), control the short-run dynamic paths towards long-run equilibrium given by equation (6.29). To test the role of the \( \gamma_{ij} \) parameters in the present value model four \( t \) tests are conducted. The first two tests are for the \( \Delta p_t \) error correction equation given by

\[
t(\gamma_{11} = 0) = \frac{0.2901 - 0}{0.0231} = 12.54, \quad t(\gamma_{12} = 0) = \frac{0.1366 - 0}{0.0841} = 1.626.
\]

The first is statistically significant whereas the second is not. Repeating these tests for the \( \Delta d_t \) error correction equation yields

\[
t(\gamma_{21} = 0) = \frac{0.0006 - 0}{0.0030} = 0.21, \quad t(\gamma_{22} = 0) = \frac{0.8792 - 0}{0.0110} = 80.17.
\]

Here the second test is statistically significant whereas the first is not. This group of tests shows that own lags are statistically important in contributing to the short-run dynamic adjustment paths, but other lagged variables are not.

The results of the weak exogeneity tests show that at the 1% significance level it is \( d_t \) that primarily adjusts to the error correction term \( u_{t-1} \), while \( p_t \) is weakly
CHAPTER 6. COINTEGRATION

exogenous. Combining this result with the transient dynamic tests in the $\Delta p_t$ error correction equation, $\Delta p_t$ appears to depend primarily on its own past history and to be independent of past movements of $d_t$. Following the discussion of causality in Chapter 4, these empirical results suggest that $d_t$ fails to Granger cause $p_t$ in which case $p_t$ may be interpreted as being both weakly and strongly exogenous. The findings also indicate that $d_t$ is the primary endogenous variable in the system and it is this variable that plays the primary role in making adjustments towards the long-run equilibrium path.

6.6 Cointegration and the Gordon Model

The present value model of equity price determination is based on a bivariate relationship between log equity prices and log dividends. An early influential paper on equity prices by Gordon (1959) actually proposes two competing models of asset prices. The first is the present value model already discussed. The second is based on earnings. For this second model, it is assumed the investor buys equity in order to obtain income per share and is indifferent as to whether returns to investment are packaged in terms of the fraction of earnings distributed as a dividend or in terms of a rise in the share’s value. Defining $p_t$ as log equity prices, $d_t$ as log dividends and $e_t$ as log earnings, these two competing models of equity prices are summarised as

\[
\begin{align*}
p_t &= \beta_{10} + \beta_{11}d_t + u_{1t}, \quad \text{[Price-Dividend long-run equation]} \\
p_t &= \beta_{20} + \beta_{21}e_t + u_{2t}, \quad \text{[Price-Earnings long-run equation]} \\
\end{align*}
\]  

where $u_{1t}$ and $u_{2t}$ are disturbance terms representing the pricing errors for the two models. This set of equations is often referred to as the Gordon model. Provided that $p_t$, $d_t$ and $e_t$ are $I(1)$ and the two disturbances are $I(0)$, the system of equations in (6.34) represents two cointegrating equations within a trivariate ($N = 3$) model. The corresponding VECM now contains two error correction terms which capture the equilibrating mechanisms that push to restore the long-run equilibrium path after transient shocks to the variables may drive the system away from the equilibrium path.

Figure 6.5 gives a 3-dimensional scatter plot of monthly observations of the variables $p_t$, $d_t$ and $e_t$, from January 1871 to September 2016 for the United States. All three variables tend to move closely around a single 1-dimensional path running through the 3-dimensional space. This near-linear path is strong visual evidence supporting the presence two cointegrating equations that together work to reduce the 3-dimensional variation. By contrast, there would be no visual evidence of cointegration if the observations were scattered evenly throughout the 3-dimensional cube, a point already made regarding the 2-dimensional scatter plot in Figure 6.2. The effect of a single cointegrating equation would be to reduce the number of dimensions by 1, in which case the observations would be attracted towards a 2-dimensional surface within
the cube. The presence of a second cointegrating equation compresses the dimension of the scatter plot further from a 2-dimensional surface towards a 1-dimensional path within the 3-dimensional cube.

Figure 6.5: Three dimensional scatter plot of the equity price index, dividends and earnings (all expressed in logs). The Gordon model implies the existence of two cointegrating vectors which is illustrated by the scatter being attracted to a line in three dimensional space. The data are monthly observations from February 1871 to September 2016.

The results of performing the Johansen cointegration test to identify the number of cointegrating equations amongst the $N = 3$ variables are presented in Table 6.4. In implementing this test the VECM in (6.15) is specified as

$$\Delta y_t = \delta + \alpha (\beta^t y_{t-1} - \beta_0) + \Gamma_1 \Delta y_{t-1} + v_t,$$  

(6.35)

with the $N = 3$ variables ordered as $y_t = (p_t, d_t, e_t)$. The dimensions of the parameters are as follows: $\delta$ is a $(3 \times 1)$ vector of constants in the VECM, $\alpha$ and $\beta$ are $(3 \times r)$ matrices of error-correction and cointegrating parameters respectively where $r$ is the number of cointegrating equations, $\beta_0$ is an $(r \times 1)$ vector of constants corresponding to the $r$ cointegrating equations, and $\Gamma_1$ is an $(N \times N)$ matrix of transient dynamic parameters serving as coefficients on $\Delta y_{t-1}$. The choice of a single additional lag in this VECM is adopted for convenience to simplify the presentation of the model. In the exercises the analysis is repeated for a lag length of $k - 1$ in the VECM chosen using information criteria applied to the corresponding VAR model.

As the dimension of the system is $N = 3$ variables, there are 3 sets of hypotheses to consider. The first null hypothesis is of no cointegration ($r = 0$) with
the alternative being there is at least 1 cointegrating equation. The test statistic is 145.6367. The 5% critical value is 29.68 resulting in a \( p \) value which is less than 0.05, thereby providing strong evidence against the null of no cointegration and in favour of the alternative hypothesis of cointegration. To identify the actual number of cointegrating equations, testing proceeds to the next set of hypotheses where the null is now 1 cointegrating equation and the alternative is that there are at least 2 cointegrating equations. The test statistic for this hypothesis is 28.7756. The 5% critical value is 15.41 resulting in a rejection at the 5% level of the null hypothesis that there is just one cointegrating equation. To determine if there is an additional cointegrating equation, testing proceeds to the third set of hypotheses where the null is that there are 2 cointegrating equations. The corresponding test statistic is 1.1765 which now provides strong evidence in favour of the null hypothesis at the 5% level since the critical value of the test is 3.76. These results support the graphical analysis in Figure 6.5 which shows visual evidence of 2 cointegrating equations.

### Table 6.4

Johansen cointegration test of the Gordon model of equity prices in (6.35) between United States equity prices, dividends and earnings. The sample period is January 1871 to September 2016.

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Eigenvalue</th>
<th>Statistic</th>
<th>5% CV</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r = 0 ): No cointegration</td>
<td>0.0647</td>
<td>145.6367</td>
<td>29.68</td>
<td>0.0001</td>
</tr>
<tr>
<td>( r = 1 : 1 ) cointegrating equation</td>
<td>0.0157</td>
<td>28.7756</td>
<td>15.41</td>
<td>0.0003</td>
</tr>
<tr>
<td>( r = 2 : 2 ) cointegrating equations</td>
<td>0.0007</td>
<td>1.1765</td>
<td>3.76</td>
<td>0.2781</td>
</tr>
</tbody>
</table>

The specification of the dividends and earnings cointegrating equations in (6.34) is based on the theoretical models proposed by Gordon (1959). However, these two equations can be rearranged to generate an alternative long-run cointegrating equation that is still perfectly consistent with the cointegrating system associated with the Gordon model. For example, setting the two equations in (6.34) equal to each other eliminates \( p_t \) as follows

\[
\beta_{10} + \beta_{1d} d_t + u_{1t} = \beta_{20} + \beta_{2e} e_t + u_{2t}.
\]

Rearranging this equation for \( d_t \) as a function of \( e_t \) gives

\[
d_{t} = \beta_{30} + \beta_{3e} e_t + u_{3t}, \quad \text{[Dividend-Earnings long-run equation]} \tag{6.36}
\]

with \( \beta_{30} = (\beta_{20} - \beta_{10})/\beta_{1d}, \beta_{3e} = \beta_{2e}/\beta_{1d} \) and \( u_{3t} = (u_{2t} - u_{1t})/\beta_{1d} \). Equation (6.36) also represents a cointegrating equation as \( d_t \) and \( e_t \) are \( I(1) \), while the disturbance term \( u_{3t} \) is a linear function of \( I(0) \) disturbances which must also be \( I(0) \). Any two of the three cointegrating equations in (6.34) and (6.36) provide a valid and observationally equivalent representation of the long-run properties underlying equity prices, dividends and earnings. To obtain a
unique representation of the cointegrated system requires a specific normalisation process, such as that given in (6.34) where each equation is normalised so that the coefficient of the price variable \( p_t \) is unity.

The Gordon model can be formulated by using the price-earnings equation in (6.34) and the dividends-earnings equation in (6.36) as the two normalized cointegrating equations to be estimated by reduced rank regression. The specification of the VECM for this version of the Gordon trivariate model containing two cointegrating equations and one lag has the explicit form

\[
\Delta p_t = \delta_1 + a_{11} u_{2t-1} + a_{12} u_{3t-1} + \gamma_{11} \Delta p_{t-1} + \gamma_{12} \Delta d_{t-1} + \gamma_{13} \Delta e_{t-1} + v_{1t}, \\
\Delta d_t = \delta_2 + a_{21} u_{2t-1} + a_{22} u_{3t-1} + \gamma_{21} \Delta p_{t-1} + \gamma_{22} \Delta d_{t-1} + \gamma_{23} \Delta e_{t-1} + v_{2t}, \\
\Delta e_t = \delta_3 + a_{31} u_{2t-1} + a_{32} u_{3t-1} + \gamma_{31} \Delta p_{t-1} + \gamma_{32} \Delta d_{t-1} + \gamma_{33} \Delta e_{t-1} + v_{3t}.
\]

(6.37)

The parameter estimates of this system are presented in Table 6.5. The long-run parameter estimates on the earnings variable in the two cointegrating equations are both numerically very close to unity. Testing the restriction \( \beta_{21} = 1 \) from the price-earnings cointegrating equation in (6.34) yields a \( t \) statistic of

\[
t(\beta_{2e} = 1) = \frac{1.0900 - 1}{0.0349} = 2.5788,
\]

which is nonetheless still statistically significant at the 5% level with a \( p \) value of 0.0099. Applying the same test to the parameter \( \beta_{31} \) in the dividend-earnings cointegrating equation in (6.36) yields a \( t \) statistic of

\[
t(\beta_{3e} = 1) = \frac{0.9170 - 1}{0.0110} = -7.5170,
\]

which also leads to a rejection of the null hypothesis at the 5% level with a \( p \) value of 0.0000.

Inspection of the error correction parameter estimates in Table 6.5 shows that equity prices adjust to the equity-earnings error correction term \( \hat{u}_{2t-1} \), but not the dividend-earnings error correction term \( \hat{u}_{3t-1} \). In fact all three variables respond when the price-earnings relationship is not in long-run equilibrium. The negative sign on \( \hat{u}_{2t-1} \) in the \( \Delta p_t \) equation of the VECM shows that when equity prices operate above (below) the price-earnings long-run relationship this results in short-run decreases (increases) in equity prices.

By contrast, the positive signs on \( \hat{u}_{2t-1} \) in the dividend and earnings error correction equations show that these variables restore equilibrium by increasing (decreasing) when equity prices are above (below) the price-earnings long-run equilibrium. Dividends and earnings also adjust towards the equilibrium path when the dividend-earnings relationship is not in equilibrium with dividends falling and earnings increasing when dividends are above their long-run value (ie when \( \hat{u}_{3t-1} > 0 \)). These error-correction mechanisms show that equity prices, dividends and earnings are all endogenous within the trivariate system and in general all operate to restore long-run equilib-
Table 6.5

Reduced-rank regression parameter estimates of the Gordon model given by the VECM in (6.37), with standard errors in parentheses. The sample period is January 1871 to September 2016. The estimated cointegrating equations are

\[ p_t = 2.7677 + 1.0900e_t + \hat{u}_{2t} \]
\[ (0.0349) \]

\[ d_t = -0.5052 + 0.9170e_t + \hat{u}_{3t} \]
\[ (0.0110) \]

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \Delta p_t )</th>
<th>( \Delta d_t )</th>
<th>( \Delta e_t )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>0.0004</td>
<td>0.0011</td>
<td>0.0003</td>
</tr>
<tr>
<td></td>
<td>(0.0013)</td>
<td>(0.0002)</td>
<td>(0.0007)</td>
</tr>
<tr>
<td>( \hat{u}_{2t-1} )</td>
<td>-0.0082</td>
<td>0.0018</td>
<td>0.0048</td>
</tr>
<tr>
<td></td>
<td>(0.0033)</td>
<td>(0.0004)</td>
<td>(0.0019)</td>
</tr>
<tr>
<td>( \hat{u}_{3t-1} )</td>
<td>0.0061</td>
<td>-0.0064</td>
<td>0.0135</td>
</tr>
<tr>
<td></td>
<td>(0.0056)</td>
<td>(0.0007)</td>
<td>(0.0033)</td>
</tr>
<tr>
<td>( \Delta p_{t-1} )</td>
<td>0.2845</td>
<td>-0.0012</td>
<td>0.0459</td>
</tr>
<tr>
<td></td>
<td>(0.0232)</td>
<td>(0.0030)</td>
<td>(0.0137)</td>
</tr>
<tr>
<td>( \Delta d_{t-1} )</td>
<td>0.0768</td>
<td>0.8328</td>
<td>0.1786</td>
</tr>
<tr>
<td></td>
<td>(0.0957)</td>
<td>(0.0123)</td>
<td>(0.0562)</td>
</tr>
<tr>
<td>( \Delta e_{t-1} )</td>
<td>0.0488</td>
<td>0.0082</td>
<td>0.7889</td>
</tr>
<tr>
<td></td>
<td>(0.0261)</td>
<td>(0.0033)</td>
<td>(0.0153)</td>
</tr>
</tbody>
</table>

Dissolution in the equity market arising from temporary shocks pushing the system into disequilibrium.

The estimated model presented in Table 6.5 is based on the price-earnings and dividend-earnings cointegrating relationships in (6.34) and (6.36). To recover the estimates of the present value cointegrating relationship between \( p_t \) and \( d_t \) note that

\[ \beta_{10} = \beta_{20} - \beta_{3e} \beta_{30} / \beta_{2e}, \quad \beta_{1d} = \beta_{2e} / \beta_{3e}, \]

which are obtained by using the expressions for \( \beta_{30} \) and \( \beta_{3e} \) given immediately below equation (6.36). Using the parameter estimates reported in Table 6.5 the price-dividend cointegrating parameter estimates are

\[ \hat{\beta}_{10} = 2.7677 + 0.9170 \times 0.5052 / 1.0900 = 3.1927, \quad \hat{\beta}_{1d} = \frac{1.0900 / 0.9170}{1.1887}. \]

These estimates would also have been achieved if the cointegrating system was expressed as in expression (6.34).
6.7 Cointegration and the Yield Curve

The expectations hypothesis of the term structure of interest rates posits that the yield on a long-maturity bond is the average of the current and expected future yields on a short maturity bond (see Chapter 2). If the long-term bond has maturity of \( m = 3 \) months with yield \( r_{3t} \) and the short-term bond is an \( m = 1 \) month bond with yield \( r_{1t} \), the investor at time \( t = 0 \) can either invest $1 for 3 months at the 3-month yield or invest the $1 in a 1-month bond at the current yield, \( r_{1t} \), and reinvest at the (expected) prevailing one month yield in months two and three. To ensure that there are no arbitrage opportunities the expectations hypothesis requires that these two strategies result in the same return, so that

\[
r_{3t} = \frac{1}{3} E_t (r_{1t} + r_{1t+1} + r_{1t+2}).
\]  

(6.38)

Equation (6.38) is rearranged to generate an expression for the spread between the long and short interest rates by subtracting \( r_{1t} \) from both sides, and adding and subtracting \( r_{1t+1}/3 \) from the right-hand side as follows

\[
r_{3t} - r_{1t} = E_t \left( -\frac{2}{3} r_{1t} + \frac{1}{3} r_{1t+1} + \frac{1}{3} r_{1t+1} - \frac{1}{3} r_{1t+1} + \frac{1}{3} r_{1t+2} \right) \\
= E_t \left( -\frac{2}{3} r_{1t} + \frac{2}{3} r_{1t+1} - \frac{1}{3} r_{1t+1} + \frac{1}{3} r_{1t+2} \right) \\
= E_t \left( \frac{2}{3} \Delta r_{1t+1} + \frac{1}{3} \Delta r_{1t+2} \right).
\]  

(6.39)

The spread between the long and the short yields is expressed as a weighted sum of future changes in the 1-month short yield. This equation can be generalised to arbitrary maturities of \( m \) months according to

\[
r_{mt} - r_{1t} = E_t \sum_{j=1}^{m-1} \left( 1 - \frac{j}{m} \right) \Delta r_{1t+j}.
\]  

(6.40)

Equation (6.40) represents a system of cointegrating equations. As bond yields are found to be \( I(1) \) from the direct application of unit root tests in Chapter 5, the right hand side of this expression must be stationary as it is a function of future changes in the short-term yield which by definition are \( I(0) \). This, in turn, implies that the spread \( r_{mt} - r_{1t} \) is stationary making \( r_{mt} \) and \( r_{1t} \) cointegrated as the spread is a linear function of two \( I(1) \) variables that becomes an \( I(0) \) variable. An important property of (6.40) is that because it holds for all \( m \) the expectations theory of the term structure predicts that there are \( m - 1 \) cointegrating long-term relationships with the short-term rate \( r_{1t} \).

Figure 6.6 contains scatter pairs of United States zero coupon bond yields from December 1946 to February 1987 for maturities of \( m = \{2, 3, 4, 5, 6, 9\} \) months. Inspection of the 5 scatter pairs in the first column between the short-term yield and the 5 longer-term yields shows strong evidence of long-run
equilibria between each of these pairs of yields. This figure also shows similar long-run relationships between other combinations of yields in the other columns. This result is a reflection of the property already established for the Gordon model in Section 6.6, namely that when there are multiple cointegrating equations in a system the cointegrating equations may be expressed in a variety of forms, whereas the space defined by the cointegrating relations is well defined and unique. For the term structure data with 6 maturities, the scatter plots in 6.6 show 15 paired linkages among the various maturities. However, the overall dimension of the linkages or cointegrated system is only 5 (such as those shown in the first column) because the remaining 10 paired linkages can be recovered by taking combinations of these chosen 5 cointegrating equations. In the empirical analysis that follows, the cointegrating equations are based on the pairs of yields presented in the first column, although undertaking the analysis on other combinations of yields would also be appropriate.

Figure 6.6: Scatter plots of the United States zero coupon bond yields of various maturities. The data are monthly for the period December 1946 to February 1987.

A formal test of the number of cointegrating equations between the 6 yields
is conducted using the Johansen cointegration test. To perform this test it is necessary to specify the VECM in (6.15). To identify the lag structure of the VECM Table 6.6 contains the results of using information criteria to determine the lag length of a VAR containing all 6 yields, expressed in levels. Information criteria are presented for lags running from $k = 0$ to $k = 8$. The Schwartz (SC) and Hannan-Quinn (HQ) information statistics both choose 2 lags whereas the Akaike information criterion chooses 6 lags. Adopting the SC and HQ chosen lag structure of $k = 2$, the appropriate lag length for the VECM model, where the lagged variables are expressed as lagged differences, is $k - 1 = 1$.

Table 6.6

<table>
<thead>
<tr>
<th>Lag</th>
<th>AIC</th>
<th>SC</th>
<th>HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-11.3622</td>
<td>-11.3096</td>
<td>-11.3415</td>
</tr>
<tr>
<td>1</td>
<td>-16.5112</td>
<td>-16.1431</td>
<td>-16.3665</td>
</tr>
<tr>
<td>2</td>
<td>-16.8283</td>
<td>-16.1447</td>
<td>-16.5595</td>
</tr>
<tr>
<td>3</td>
<td>-16.8407</td>
<td>-15.8415</td>
<td>-16.4478</td>
</tr>
<tr>
<td>4</td>
<td>-16.8513</td>
<td>-15.5366</td>
<td>-16.3343</td>
</tr>
<tr>
<td>5</td>
<td>-16.9090</td>
<td>-15.2787</td>
<td>-16.2679</td>
</tr>
<tr>
<td>6</td>
<td>-16.9097</td>
<td>-14.9639</td>
<td>-16.1445</td>
</tr>
<tr>
<td>7</td>
<td>-16.8908</td>
<td>-14.6295</td>
<td>-16.0015</td>
</tr>
<tr>
<td>8</td>
<td>-16.8898</td>
<td>-14.3129</td>
<td>-15.8764</td>
</tr>
</tbody>
</table>

Given the results of the lag structure tests on the VAR the VECM in (6.15) is specified for the $N = 6$ bond yields as

$$\Delta y_t = \alpha (\beta' y_{t-1} - \beta_0) + \Gamma_1 \Delta y_{t-1} + v_t,$$  (6.41)

with the yields ordered as $y_t = (r_{9t}, r_{6t}, r_{5t}, r_{4t}, r_{3t}, r_{2t})$, where the subscript gives the maturity of the bond. The respective error-correction and cointegrating parameter matrices $\alpha$ and $\beta$ are $(6 \times r)$ where $r$ is the number of cointegrating equations which, from the theory of the term structure of interest rates, is $r = N - 1 = 5$. The $(r \times 1)$ vector of constants $\beta_0$ allows for the possibility of non-zero risk premia in the longer-term bond yields, while $\Gamma_1$ is a $(6 \times 6)$ matrix of short-run dynamic parameters. This choice of specification is based also on two additional restrictions regarding deterministic time trends. The first is that there are no time trends in the cointegrating equations, $\phi = 0$ in equation (6.15), as the risk premia on bond yields are assumed constant given that spreads are not observed to widen persistently over time. The second is that there are no intercepts in the error correction equations, $\delta = 0$ in equation (6.15). This set of restrictions is different to the VECM specified for the Gordon model and is motivated by the property that as bonds have a
finite maturity, bond yields cannot increase or decrease without bound over
time, as would be the case with series exhibiting deterministic trends.

The results of the Johansen cointegration test are given in Table (6.7) based on
the VECM in (6.41). There are \( N = 6 \) sets of hypotheses, where the first null
is of no cointegration amongst the \( I(1) \) yields, and the last null corresponds to
\( N − 1 = 5 \) cointegrating equations. The first 5 null hypotheses are all soundly
rejected at the 5% level with \( p \) values of 0.0001 or smaller. The last null hy-
pothesis of 5 cointegrating equations is not rejected at the 5% level given the
\( p \) value is \( 0.3335 > 0.05 \). This sequence of tests supports the visual analysis of
the scatter plots presented in Figure 6.6 showing strong evidence of 5 cointe-
grating equations amongst the 6 bond yields in \( y_t \).

Table 6.7

<table>
<thead>
<tr>
<th>Null Hypothesis</th>
<th>Eigenvalue</th>
<th>Statistic</th>
<th>5% CV</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r = 0 ): No cointegration</td>
<td>0.3023</td>
<td>557.1008</td>
<td>103.8473</td>
<td>0.0000</td>
</tr>
<tr>
<td>( r = 1 ): 1 cointegrating equation</td>
<td>0.2421</td>
<td>383.9215</td>
<td>76.9728</td>
<td>0.0001</td>
</tr>
<tr>
<td>( r = 2 ): 2 cointegrating equations</td>
<td>0.2228</td>
<td>250.5879</td>
<td>54.0790</td>
<td>0.0000</td>
</tr>
<tr>
<td>( r = 3 ): 3 cointegrating equations</td>
<td>0.1448</td>
<td>129.3162</td>
<td>35.1927</td>
<td>0.0000</td>
</tr>
<tr>
<td>( r = 4 ): 4 cointegrating equations</td>
<td>0.0978</td>
<td>54.0827</td>
<td>20.2618</td>
<td>0.0000</td>
</tr>
<tr>
<td>( r = 5 ): 5 cointegrating equations</td>
<td>0.0095</td>
<td>4.5730</td>
<td>9.1645</td>
<td>0.3335</td>
</tr>
</tbody>
</table>

The estimates of the long-run equations based on the VECM in (6.41) with
\( r = 5 \) cointegrating equations are

\[
\begin{align*}
  r_{9t} &= 0.2450 + 1.0424 r_{2t} + \tilde{u}_{1t}, \\
  r_{6t} &= 0.1662 + 1.0364 r_{2t} + \tilde{u}_{2t}, \\
  r_{5t} &= 0.1379 + 1.0301 r_{2t} + \tilde{u}_{3t}, \\
  r_{4t} &= 0.0984 + 1.0225 r_{2t} + \tilde{u}_{4t}, \\
  r_{3t} &= 0.0518 + 1.0133 r_{2t} + \tilde{u}_{5t}.
\end{align*}
\]

(6.42)

All of the long-run slope parameter estimates are numerically close to the the-
oretical value of unity, although the differences are statistically significant
based on \( t \) tests. For example, in the case of the cointegrating equation be-
tween \( r_{9t} \) and \( r_{2t} \), the \( t \) statistic is

\[
t(\beta_{12} = 1) = \frac{1.0424 - 1.0}{0.0134} = 3.1642,
\]

\[
(0.0134) 
\]

\[
(0.0134)
\]
which has a \( p \) value of 0.0000 providing strong rejection of the null hypothesis that \( \beta_{12} = 1 \).

Inspection of the parameter estimates on the intercepts in (6.42) reveals positive risk premia which increase with maturity starting from \( \hat{\beta}_{50} = 0.0518 \) for the 3-month yield and growing to \( \hat{\beta}_{10} = 0.2450 \) for the 9-month yield. The risk premia are also found to be statistically significant for all spreads. In the case of the spread between \( r_{9t} \) and \( r_{2t} \), the \( t \) statistic associated with the null hypothesis that \( \beta = 0 \), is

\[
 t(\beta = 0) = \frac{0.2450 - 0.0}{0.0809} = 3.03.
\]

The \( p \) value is 0.0000 which provides strong evidence against the null hypothesis of no risk premia. All of the statistical tests show that whilst there is strong evidence in favour of the term structure of interest rates model in (6.40), the statistical evidence does reject the ‘pure’ form of this model with \((1, -1)\) spreads and no risk premia.

To investigate the stability properties of the estimated model of the term structure, Figure 6.7 shows the impulse responses of all 6 bond yields to a shock in the 2-month yield as given by the last element in \( v_t \), namely \( v_{6t} \), in (6.41), namely \( v_{2t} \).

The impulse responses are shown for a 24-month or 2-year horizon. The system is stable with all impulse responses nearly converging to a steady state level within a year, thereby providing strong support for cointegration and the presence of long-run equilibria among the 6 yields. The impulse responses after 2 years reported in Figure 6.7 are

\[
\frac{\partial y_{1t+24}}{\partial v_{6t}} = 5065, \quad \frac{\partial y_{2t+24}}{\partial v_{6t}} = 0.5038, \quad \frac{\partial y_{3t+24}}{\partial v_{6t}} = 0.5007, \\
\frac{\partial y_{4t+24}}{\partial v_{6t}} = 0.4971, \quad \frac{\partial y_{5t+24}}{\partial v_{6t}} = 0.4928, \quad \frac{\partial y_{6t+24}}{\partial v_{6t}} = 0.4864.
\]

These estimates can also be used to generate the long-run relationships between the yields which provides an internal consistency check on the cointegrating parameter estimates in (6.42). In the case of the long yield, \( y_{1t} \), and the short yield, \( y_{6t} \), this relationship is

\[
\frac{\partial r_{9t+24}}{\partial r_{2t+24}} = \frac{\partial y_{1t+24}}{\partial y_{6t+24}} = \frac{\partial y_{1t+24}}{\partial v_{6t}} \frac{\partial v_{6t}}{\partial y_{6t+24}} = \frac{0.5065}{0.4864} = 1.0413,
\]

which is in agreement with the long-run estimate of 1.0424 reported for the first long-run equation in (6.42). Extending the impulse horizon to 3 years

\[12\] Technically the size of the shock to the 2-month yield is not determined by the standard deviation of \( v_{6t} \) in the VECM in equation (6.41), as the impulse responses are expressed in terms of structural shocks using the Cholesky triangular decomposition, discussed in Chapter 4, with the ordering running from the 2-month to the 9-month yield.
improves the accuracy of this calculation to 4 decimal places, and in the limit the estimate obtained from the impulse responses would perfectly match the estimate obtained from the cointegrating parameter estimate.

6.8 Exercises

The data required for the exercises are available for download as EViews workfiles (*.wf1), Stata datafiles (*.dta), comma delimited text files (*.csv) and as Excel spreadsheets (*.xlsx).

1. The Present Value Model
The data are monthly observations on United States equity prices and dividends for the period January 1871 to September 2016. The present value model predicts the following relationship between price and dividends

\[ p_t = \beta_0 + \beta_d d_t + u_t, \]

where \( p_t \) is the log equity price, \( d_t \) is the log of dividend payments, \( u_t \) is a disturbance term, \( \beta_0 = -\log \delta \), \( \delta \) is the discount rate, and \( \beta_d = 1 \) if the present value model holds.

(a) Estimate a bivariate VAR with a fitted intercept for the variables \( p_t \) and \( d_t \), choosing an appropriate lag structure \( k \).

(b) Test for cointegration between \( y_t = (p_t, d_t) \) by specifying the VECM

\[ \Delta y_t = \delta + \alpha (\beta' y_{t-1} - \beta_0) + \sum_{j=1}^{k-1} \Gamma_j \Delta y_{t-j} + v_t, \]

where the number of lags in the VECM is determined by the choice of the lag structure of the VAR in part (a).

(c) Given the results in part (b) estimate a bivariate ECM for \( p_t \) and \( d_t \). Interpret the results paying particular attention to the long-run parameter estimates, \( \hat{\beta}_0 \) and \( \hat{\beta}_d \), and the error correction parameter estimates, \( \hat{\alpha}_1 \) and \( \hat{\alpha}_2 \).

(d) Derive an estimate of the long-run real discount rate from \( \exp(-\beta_0) \) and interpret the result.

(e) Test the restriction \( H_0 : \beta_d = 1 \). Discuss whether the empirical results support the present value model.

2. Forward Market Efficiency

The data files contain weekly data (all recorded on a Wednesday) for the period 4 January 1984 to 31 December 1990 on the spot $/AUD exchange rate together with forward rates for 1, 3 and 6 months. The data are from Corbae, Lim and Ouliaris (1992) who test for speculative efficiency by considering the equation

\[ s_t = \beta_0 + \beta_f f_{t-n} + u_t, \]

where \( s_t \) is the log spot rate, \( f_{t-n} \) is the log forward rate lagged \( n \) periods and \( u_t \) is a disturbance term. In the case of weekly data and the forward rate is the 1-month rate, \( f_{t-4} \) is an unbiased estimator of \( s_t \) if \( \beta_f = 1 \).
(a) Use unit root tests to determine the level of integration of $s_t$ and $f_t$.

(b) Test for cointegration between $y_t = (s_t, f_{t-4})$ by specifying the VECM
\[ \Delta y_t = \alpha (\beta' y_{t-1} - \beta_0) + v_t. \]

(c) Given the results in part (b), estimate a bivariate VECM.

(d) Interpret the parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_f$, and test the restriction $\beta_f = 1$.

(e) Repeat parts (a) to (d) for the 3 month and 6 month forward rates.
*Hint: remember that the frequency of the data is weekly.*

3. Fisher Hypothesis

The data are quarterly observations on United States inflation and interest rates from 1954:Q3 to 2008:Q4. Under the Fisher hypothesis the nominal interest rate $i_t$ fully reflects the long-run movements in the inflation rate $\pi_t$. The Fisher hypothesis is represented by
\[ i_t = \beta_0 + \beta_\pi \pi_t + u_t, \]
where $u_t$ is a disturbance term and the slope parameter is $\beta_\pi = 1$ if the Fisher hypothesis holds.

(a) Construct the percentage annualised inflation rate, $\pi_t$.

(b) Use a bivariate VAR (with an intercept) for $i_t$ and $\pi_t$ to choose an appropriate lag structure $k$.

(c) Test for cointegration between $i_t$ and $\pi_t$) based on the VECM
\[ \Delta y_t = \alpha (\beta' y_{t-1} - \beta_0) + \sum_{j=1}^{k-1} \Gamma_j \Delta y_{t-j} + v_t, \]
where and the number of lags in the VECM is determined by the optimal lag structure of the VAR chosen in part (b).

(d) Test for cointegration subject to the restriction that $\beta_\pi = 1$. Does the Fisher hypothesis hold in the long-run? Discuss.

4. Purchasing Power Parity
The data are monthly observations for the period January 1979 to December 2008 for the consumer price index in Australia and the United States, respectively, together with the $/AUD exchange rate. Under the assumption of purchasing power parity (PPP), the nominal exchange rate, $S$, adjusts in the long-run to the price differential between foreign ($F$) and domestic ($P$) countries according to

$$S = \frac{P}{F}.$$ 

This expression suggests that the relationship between the nominal exchange rate and the prices in the two countries is given by

$$s_t = \beta_0 + \beta_P p_t + \beta_f f_t + u_t,$$

where lower case letters denote variables expressed in logarithms and $u_t$ is a disturbance term representing short-run departures from PPP.

(a) Construct the relevant variables, $s_t$, $f_t$, $p_t$ and the foreign price differential $p_t - f_t$.

(b) Use unit root tests to determine the level of integration of all of these series based on 12 lags. Discuss the results in terms of the level of integration of each series.

(c) Test for cointegration between $y_t = (s_t, p_t, f_t)$ using the VECM

$$\Delta y_t = \delta + \alpha(\beta' y_{t-1} - \beta_0) + \sum_{j=1}^{k-1} \Gamma_j \Delta y_{t-j} + v_t$$

with $k = 12$ lags.

(d) Given the results in part (c) estimate a trivariate VECM for $s_t$, $p_t$ and $f_t$.

(e) Interpret the long-run parameter, error correction and short-run parameter estimates.

(f) Test the restriction $H_0 : \beta_p = -\beta_f$. Is PPP satisfied?

5. The Gordon Equity Model
The data are monthly observations on United States equity prices, dividends and earnings for the period January 1871 to September 2016. The Gordon equity model (1959), is represented by two long-run equations involving the variables log equity prices \( p_t \), log dividends \( d_t \), and log earnings \( e_t \), according to the specification

\[
\begin{align*}
    p_t &= \beta_{10} + \beta_{11}d_t + u_{1t}, \\
    p_t &= \beta_{20} + \beta_{21}e_t + u_{2t},
\end{align*}
\]

where \( u_{1t} \) and \( u_{2t} \) are disturbance terms representing the pricing errors for the two models.

(a) Estimate a VAR (with an intercept) containing \( p_t, d_t \) and \( e_t \) based on an optimal choice of the lag structure \( k \).

(b) Test for cointegration between \( y_t = (p_t, d_t, e_t) \) by specifying the VECM

\[
\Delta y_t = \delta + \alpha (\beta' y_{t-1} - \beta_0) + \sum_{j=1}^{k-1} \Gamma_j \Delta y_{t-j} + v_t,
\]

where the number of lags in the VECM is determined by the optimal lag structure of the VAR obtained in part (a).

(c) Given the results in part (b) estimate the VECM and write out the estimated long-run cointegrating equation(s) in terms of earnings, \( e_t \). Interpret the long-run parameter estimates.

(d) Derive estimates of the long-run relationship between \( p_t \) and \( d_t \) by

i. Using the parameter estimates reported in part (c).

ii. Re-estimating the VECM with the ordering of the variables chosen as \( y_t = (p_t, e_t, d_t) \).

(e) Examine the stability of the estimated VECM and use the long-run estimates of the impulse responses to recompute the long-run cointegrating parameter estimates given in part (c).

6. The Term Structure of Interest Rates

The data are monthly observations from December 1946 to February 1987 on United States zero coupon bond yields for maturities ranging from 2 months to 9 months.

The expectations hypothesis of the term structure of interest rates predicts the following relationship between a long-term interest rate of maturity \( n \) and a short-term rate of maturity \( m < n \)

\[
r_{nt} = \beta_0 + \beta_mr_{mt} + u_t,
\]
where $u_t$ is a disturbance term and $\beta_0$ represents the term premium. The pure expectations hypothesis requires that $\beta_0 = 0$ and $\beta_m = 1$.

(a) Estimate a VAR with a fitted intercept and $N = 6$ variables containing the yields $y_t = (r_{9t}, r_{6t}, r_{5t}, r_{4t}, r_{3t}, r_{2t})$ using a data-determined lag structure $k$.

(b) Test for cointegration amongst the $N = 6$ yields by specifying the VECM

$$\Delta y_t = \alpha (\beta' y_{t-1} - \beta_0) + \sum_{j=1}^{k-1} \Gamma_j \Delta y_{t-j} + v_t,$$

where the number of lags in the VECM is based on the optimal lag structure of the VAR from part (a).

(c) Given the results in part (b), estimate the VECM and write out the estimated long-run equations in terms of the short-term yield $r_{2t}$.

(d) Perform a joint test that all of the cointegrating parameters are of the form $(1, -1)$.

(e) Test each of the $N = 6$ yields individually for weak exogeneity.

(f) Given the results in part (e), test each of the $N = 6$ yields for strong exogeneity.

7. Spurious Regression Problem

If there is no long-run relationship binding $I(1)$ variables over time, a regression involving such unrelated or independent nonstationary variables should not be expected to find any statistical relationship connecting them, as is the case when the variables are independent and stationary. Spurious regressions among $I(1)$ variables were investigated in a simulation study by Granger and Newbold (1974), showing that major inferential distortions can arise in such cases. Phillips (1986; 1998) demonstrates that in regressions of this type the $t$ statistics will always diverge to $+\infty$ because the trends within each series inevitably correlate with each other and suggest statistical significance. Against this background, the present exercise is based on a Monte Carlo study conducted by Banerjee, Dolado, Galbraith and Hendry (1993) that investigates the spurious regression problem.

(a) Consider the following bivariate models

\begin{align*}
(i) & \quad y_{1t} = v_{1t}, \
(ii) & \quad y_{1t} = y_{1t-1} + v_{1t}, \
(iii) & \quad y_{1t} = y_{1t-1} + v_{1,t}, \
(iv) & \quad y_{1t} = 2y_{1t-1} - y_{1t-2} + v_{1t},
\end{align*}

\begin{align*}
(i) & \quad y_{2t} = v_{2t}, \
(ii) & \quad y_{2t} = y_{2t-1} + v_{2t}, \
(iii) & \quad y_{2t} = 2y_{2t-1} - y_{2t-2} + v_{2t}, \
(iv) & \quad y_{2t} = 2y_{2t-1} - y_{2t-2} + v_{2t},
\end{align*}

in which $v_{1t}$ and $v_{2t}$ are both $N(0, 1)$. Simulate each bivariate model 10000 times for a sample of size $T = 100$ and compute the correlation coefficient, $\hat{\rho}$, of each draw. Compute the sampling distributions of $\hat{\rho}$ for the four sets of bivariate models and discuss the
properties of these distributions in the context of the spurious regression problem.

(b) Repeat part (a) with $T = 500$. What do you conclude?

(c) Repeat part (a), except for each draw estimate the following regression equation by least squares

$$y_{2t} = \beta_0 + \beta_1 y_{1t} + u_t.$$ 

Compute the sampling distributions of the least squares estimator $\hat{\beta}_1$ and its $t$ statistic for the four sets of bivariate models. Discuss the properties of these distributions in the context of the spurious regression problem.
Chapter 7

Forecasting

Considerations of the future values of financial and economic variables play an important role in decision making by agents in financial markets. Forecasting the future course of financial markets and economic activity is therefore of intense interest to investors, banking institutions, policy makers, and regulatory authorities.

A forecast is a quantitative estimate about the most likely future value of a particular variable. Forecasts are typically based on past and current information about the variable itself and other observable variables that are thought to be related to it. In econometric forecasting, this information is typically embodied in an empirical model whose solution shows the dependence of the variable of interest on other observable variables and unobserved random variables representing errors and disturbances. The mechanism of econometric forecasting then relies on the estimation of the equations of these models with observed data and the use of the fitted equations to create projections of future values.

Previous chapters have studied a wide variety of econometric models suited to financial time series data, covering both univariate and multivariate models. The empirical specification and estimation of these models provides a groundwork for producing forecasts that are objective in the sense that they can be replicated exactly with knowledge of the structure of the model, the estimation procedure, and the data used in estimation. Replicability is important because it provides a mechanism for evaluating the performance characteristics of a particular model or method in relation to other models and methods. This approach contrasts with methods that involve purely subjective assessments that are not reproducible or testable in simulation exercises.

Forecasting also serves a useful purpose as a means to compare and rank alternative models and to assess different methods of estimation. In carrying out such exercises, forecast errors are useful in directing attention towards the
potential weaknesses in model specification that lead to systematic errors in forecast performance. Forecast evaluation based on past successes and failures also provides a useful way of choosing between alternative models and a way to combine models using the information that is contained in past performance.

7.1 Types of Forecasts

Illustrative examples of forecasting in financial markets are as follows.

(i) The determination of the price of an asset based on present value methods requires discounting the present and forecasted future dividend stream at a discount rate that may be allowed to change over time.

(ii) Firms are interested in forecasting the future health of the economy when making decisions about current capital outlays. Investments in capital equipment earn streams of returns over time that depend on the state of the economy and that need to be discounted to assess the viability of the investments.

(iii) In currency markets, forward exchange rates may be used to provide an estimate, or forecast, of the future spot exchange rate.

(iv) In options markets, the Black-Scholes method and other methods for pricing options rely on present information such as the price of the underlying asset, the strike price, and expiration date together with forecasts (or assumptions) about the asset’s volatility over the life of the option.

(v) In futures markets, buyers and sellers enter a contract to buy and sell commodities at a future date based on forecasts of future prices of those commodities.

(vi) Model-based computation of Value-at-Risk requires repeated forecasting of the value of a portfolio over a given time horizon.

Although these illustrations differ considerably, the principles and issues involved in forecasting are typically similar, involving the form of the model being used, the nature of the data, and the methods of projection employed. Before developing the mechanisms for forecast generation some terminology is useful.

Consider an observed sample of data \( \{y_1, y_2, \ldots, y_T\} \) and an econometric model that is to be used to generate forecasts of \( y \) over a horizon of \( H \) periods. The forecasts of \( y \) are denoted by \( \hat{y} \) and are of two main types.
7.1. TYPES OF FORECASTS

Ex Ante Forecasts: The entire sample \( \{y_1, y_2, \ldots, y_T\} \) is used to estimate the model and the task is to forecast the variable \( y \) over the future horizon \( T + 1 \) to \( T + H \).

Ex Post Forecasts: The model is estimated over a restricted sample period that excludes the last \( H \) observations, \( \{y_1, y_2, \ldots, y_{T-H}\} \). The model is then forecasted out-of-sample from \( y_{T-H+1} \) through to \( y_T \). Since the actual values of these later observations are known it is possible to compare the accuracy of the forecasts with the actual values.

Ex post and ex ante forecasts may be illustrated as follows:

<table>
<thead>
<tr>
<th>Sample</th>
<th>( y_1, y_2, \ldots, y_{T-H}, y_{T-H+1}, y_{T-H+2}, \ldots, y_T )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ex Post</td>
<td>( y_1, y_2, \ldots, y_{T-H}, \hat{y}<em>{T-H+1}, \hat{y}</em>{T-H+2}, \ldots, \hat{y}_{T} )</td>
</tr>
<tr>
<td>Ex Ante</td>
<td>( y_1, y_2, \ldots, y_{T-H}, y_{T-H+1}, y_{T-H+2}, \ldots, y_T, \hat{y}<em>{T+1}, \ldots, \hat{y}</em>{T+H} )</td>
</tr>
</tbody>
</table>

It is clear therefore that forecasting ex ante for \( H \) periods ahead requires the successive generation of \( \hat{y}_{T+1}, \hat{y}_{T+2} \) up to and including \( \hat{y}_{T+H} \). This is referred to as a multi-step dynamic forecast. On the other hand, ex post forecasting allows some latitude for choice. The forecast \( \hat{y}_{T-H+1} \) is based on data up to and including \( y_{T-H} \). In generating the forecast \( \hat{y}_{T-H+2} \) the observation \( y_{T-H+1} \) is available for use. Forecasts that use this observation are referred to as one-step ahead or static forecasts. Ex post forecasting also allows for multi-step forecasting using data up to and including \( y_{T-H} \), known as dynamic forecasting.

There is a distinction between forecasting based on dynamic time series models and forecasts based on broader linear or nonlinear regression models. Forecasts based on dynamic univariate or multivariate time series models such as those developed in Chapter 4 are sometimes referred to as recursive forecasts. Forecasts that are based on structural econometric models are sometimes known as structural forecasts. However, the distinction between these two types of forecasts is often unclear because econometric models often involve both structural and dynamic time series features.

Finally, forecasts in which only a single quantity \( \hat{y}_{T+H} \) is reported for period \( T + H \) are known as point forecasts. A point forecast of \( y_{T+H} \) represents an estimate of this future value of \( y \). Even if it is known from past performance that this estimate is a particularly good one, there is inevitably uncertainty associated with every such forecast. Interval forecasts represent this uncertainty by providing a range of forecast values about the estimate \( \hat{y}_{T+H} \), called a prediction interval, within which the actual value \( y_{T+H} \) is expected to lie with some given level of confidence. Density forecasting goes beyond interval forecasting by building an estimate of the probability distribution of a future value of \( y \) conditional on past information. These density forecasts can be particularly useful in financial risk management where value at risk calculations are needed which, in turn, depend on the probability density forecasts.
of portfolio values.

7.2 Forecasting Univariate Time Series Models

To understand the most basic principles of forecasting from financial econometric models a univariate AR(1) model is sufficient to demonstrate the key elements. Extending the model to more general univariate and multivariate models only increases the complexity to the computation but not the underlying techniques of how the forecasts are generated.

AR(1) Model

Consider the AR(1) model

\[ y_t = \phi_0 + \phi_1 y_{t-1} + v_t, \quad v_t \sim iid N(0, \sigma_v^2). \]  \(7.1\)

Suppose that the data consist of \( T \) sample observations \( y_1, y_2, \ldots, y_T \). Now consider using the model to forecast the variable one period into the future, at \( T + 1 \). If there is no change in the generating mechanism the model at time \( T + 1 \) is

\[ y_{T+1} = \phi_0 + \phi_1 y_T + v_{T+1}. \]  \(7.2\)

To be able to compute a perfect forecast of \( y_{T+1} \) it is necessary to know everything on the right-hand side of equation \(7.2\). Inspection of this equation reveals that some of these terms are known and some are unknown at time \( T \):

- **Observations:** \( y_T \) Known,
- **Parameters:** \( \phi_0, \phi_1 \) Unknown,
- **Disturbance:** \( v_{T+1} \) Unknown.

The aim of forecasting is to produce the best possible estimate of \( y_{T+1} \). A natural approach is then to replace the unknowns above with their best estimates. In the case of the parameters the obvious procedure is to replace them with point estimates, \( \hat{\phi}_0 \) and \( \hat{\phi}_1 \), that are known to have good properties such as consistency and where the full sample is used to obtain the estimates. Formally this involves replacing the conditional mean \( \phi_0 + \phi_1 y_T \) of the distribution of \( y_{T+1} \) by the sample estimate \( \hat{\phi}_0 + \hat{\phi}_1 y_T \). If the estimates \( (\hat{\phi}_0, \hat{\phi}_1) \) are consistent then the conditional mean estimate \( \hat{\phi}_0 + \hat{\phi}_1 y_T \) will be consistent for \( \phi_0 + \phi_1 y_T \). In the same way, the unknown disturbance term \( v_{T+1} \) in \(7.2\) is replaced by using the mean of its distribution, which will be simply \( E(v_{T+1}) = 0 \) if the model is correctly specified and there is no structural change in the forecast period. The resulting forecast of \( y_{T+1} \) based on equation \(7.2\) is given by

\[ \hat{y}_{T+1} = \hat{\phi}_0 + \hat{\phi}_1 y_T + 0 = \hat{\phi}_0 + \hat{\phi}_1 y_T, \]  \(7.3\)
7.2. FORECASTING UNIVARIATE TIME SERIES MODELS

where \( \hat{y}_{T+1} \) signifies that it is a forecast quantity.

Now consider extending the forecast horizon to \( T + 2 \), the second period after the end of the sample period. The strategy is the same as before in which the first step is to express the model at time \( T + 2 \) as

\[
y_{T+2} = \phi_0 + \phi_1 y_{T+1} + \nu_{T+2},
\]

in which all terms are now unknown at the end of the observation period at time \( T \), that is:

- Observations: \( y_{T+1} \) Unknown,
- Parameters: \( \phi_0, \phi_1 \) Unknown,
- Disturbance: \( \nu_{T+2} \) Unknown.

As before, the parameters \( \phi_0 \) and \( \phi_1 \) are replaced by the estimates \( \hat{\phi}_0 \) and \( \hat{\phi}_1 \), respectively, and the disturbance \( \nu_{T+2} \) by its mean \( \text{E}[\nu_{T+2}] = 0 \). What is new in equation (7.4) is the presence of the unknown quantity \( y_{T+1} \) on the right-hand side of the equation. Again, the strategy of replacing unknowns by a best estimate suggests that the forecast of this variable obtained in the previous step, \( \hat{y}_{T+1} \), be used in place of \( y_{T+1} \). Accordingly, the forecast for the second period is obtained

\[
\hat{y}_{T+2} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+1} + 0 = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+1}.
\]

Clearly extending this analysis to \( H \) periods ahead implies a forecasting equation of the form

\[
\hat{y}_{T+H} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+H-1} + 0 = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+H-1}.
\]

The need to use the result from the previous step to generate a forecast in the subsequent step is commonly referred to as recursive forecasting. The process is dynamic because earlier forecasts are needed in order to produce forecasts at longer horizons. Moreover, as all of the information embedded in the forecasts \( \hat{y}_{T+1}, \hat{y}_{T+2}, \ldots, \hat{y}_{T+H} \) is based on information up to and including the last observation in the sample at time \( T \), the forecasts are commonly referred to as conditional mean forecasts where conditioning is based on information available at the end of the observation period which is time \( T \).

**Forecast Properties**

Looking at recursive forecasts as successive estimates of corresponding conditional means is useful conceptually and in developing statistical properties of these forecasts. To begin, consider the expected value of the next period observation given information to time \( T \), which is given by

\[
\text{E}_T(y_{T+1}) = \phi_0 + \phi_1 y_T.
\]
Under quite general conditions, least squares estimates \((\hat{\phi}_0, \hat{\phi}_1)\) are consistent estimates of \((\phi_0, \phi_1)\), and so the conditional expectation is consistently estimated by the forecast \(\hat{y}_{T+1}\) in (7.3). In the same way, the conditional expectation of \(y_{T+2}\) given information to time \(T\) is

\[
E_T(y_{T+2}) = E_T(\phi_0 + \phi_1 y_{T+1} + v_{t+2}) = \phi_0 + \phi_1 (\phi_0 + \phi_1 y_T) = \phi_0 + \phi_0 \phi_1 + \phi_1^2 y_T.
\]

By analogy it follows that

\[
E_T(y_{T+3}) = \phi_0 + \phi_0 \phi_1 + \phi_0 \phi_1^2 + \phi_1^3 y_T,
\]

and more generally at horizon \(H\),

\[
E_T(y_{T+h}) = \phi_0 + \phi_0 \phi_1 + \phi_0 \phi_1^2 + \cdots + \phi_0 \phi_1^{H-1} + \phi_1^H y_T = \phi_0 (1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^{H-1}) + \phi_1^H y_T. \tag{7.5}
\]

This conditional expectation is consistently estimated by the forecast

\[
\hat{y}_{T+H} = \hat{\phi}_0 (1 + \hat{\phi}_1 + \hat{\phi}_1^2 + \cdots + \hat{\phi}_1^{H-1}) + \hat{\phi}_1^H y_T. \tag{7.6}
\]

Using (7.5) and (7.6) a number of key properties of these recursive forecasts of conditional expectations may be obtained.

(i) When the model is stationary with \(|\phi_1| < 1\), then it follows from (7.5) that as \(H \to \infty\)

\[
E_T(y_{T+H}) \to \frac{\phi_0}{1 - \phi_1}.
\]

In other words, the conditional mean converges to the unconditional mean of \(y_t\). This result relies on \((1 + \phi_1 + \phi_1^2 + \cdots + \phi_1^{H-1})\) being a convergent series and also on the elimination of the second term on the right-hand side of (7.5). Both these conditions follow from \(|\phi_1| < 1\) by virtue of the stationarity assumption. Further, when \((\hat{\phi}_0, \hat{\phi}_1)\) is consistent for \((\phi_0, \phi_1)\), the forecast \(\hat{y}_{T+H}\) in (7.6) converges to the unconditional mean, \(\phi_0 / (1 - \phi_1)\), as the sample size \(T \to \infty\) and the horizon \(H \to \infty\).

(ii) The forecasts at all horizons are also consistent estimates of the corresponding conditional means. The conditional mean may be regarded as an optimal forecast because it does not involve any variability due to the estimation of unknown parameters. The optimal forecast errors at differing horizons may be obtained from (7.5) as follows

\[
y_{T+1} - E_T(y_{T+1}) = \phi_0 + \phi_1 y_T + v_{t+1} - (\phi_0 + \phi_1 y_T) = v_{t+1}
\]

\[
y_{T+2} - E_T(y_{T+2}) = \phi_0 + \phi_0 \phi_1 + \phi_1^2 y_T + v_{t+2} + \phi_1 v_{t+1} - (\phi_0 + \phi_0 \phi_1 + \phi_1^2 y_T)
\]

\[= v_{t+2} + \phi_1 v_{t+1}.
\]
7.2. FORECASTING UNIVARIATE TIME SERIES MODELS

It follows that the optimal forecast error at $T + H$ can be written as
\[
y_{T+H} - E_T(y_{T+H}) = v_{t+H} + \phi_1 v_{t+H-1} + \cdots + \phi^{H-1} v_{t+1}. \tag{7.7}
\]

Since $E(v_{t+H}) = 0$ for all $H$ the expected value of the optimal forecast error is 0 and hence the optimal forecast is unbiased.

(iii) The results established so far imply that the variance of the forecasts at different horizons are as follows
\[
\text{var}[y_{T+1} - E_T(y_{T+1})] = \sigma^2_v
\]
\[
\text{var}[y_{T+2} - E_T(y_{T+2})] = \sigma^2_v(1 + \phi_1^2)
\]
\[= \vdots
\]
\[
\text{var}[y_{T+H} - E_T(y_{T+H})] = \sigma^2_v(1 + \phi_1^2 + \phi_1^4 + \phi_1^6 + \cdots + \phi_1^{2(H-1)}),
\]
where $\sigma^2_v = E(v^2_t)$ is the variance of the one-step-ahead optimal forecast error. Importantly, the variance of the optimal forecast is an increasing function of the forecast horizon $H$. Thus, even in the optimal situation where no parameters need to be estimated, raising the forecast horizon inevitably raises the forecast error variance.

AR(2) Model

Extending the AR(1) model to an AR(2) model gives
\[
y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + v_t,
\]
and the same simple strategy may be used to forecast $y_t$ in this model. First, the model at time $T + 1$ is written as
\[
y_{T+1} = \phi_0 + \phi_1 y_T + \phi_2 y_{T-1} + v_{T+1}.
\]
Replacing the parameters $\{\phi_0, \phi_1, \phi_2\}$ by their sample estimates $\{\hat{\phi}_0, \hat{\phi}_1, \hat{\phi}_2\}$ and the disturbance $v_{T+1}$ by its mean $E[v_{T+1}] = 0$, the forecast for the first period into the future is
\[
\hat{y}_{T+1} = \hat{\phi}_0 + \hat{\phi}_1 y_T + \hat{\phi}_2 y_{T-1}.
\]
To generate the forecasts for the second period, the AR(2) model is written at time $T + 2$ as
\[
y_{T+2} = \phi_0 + \phi_1 y_{T+1} + \phi_2 y_T + v_{T+2}.
\]
Replacing all of the unknowns on the right-hand side by their corresponding estimates, including the forecast $\hat{y}_{T+1}$ of $y_{T+1}$, gives
\[
\hat{y}_{T+2} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+1} + \hat{\phi}_2 y_T.
Similarly, to derive the forecast of \( y_t \) at time \( T + 3 \) the AR(2) model is written at \( T + 3 \) as

\[
y_{T+3} = \phi_0 + \phi_1 y_{T+2} + \phi_2 y_{T+1} + \nu_{T+3}.
\]

All terms on the right-hand side are unknown and need to be estimated, so the forecasting equation becomes

\[
\hat{y}_{T+3} = \hat{\phi}_0 + \hat{\phi}_1 \hat{y}_{T+2} + \hat{\phi}_2 \hat{y}_{T+1}.
\]

The same procedure applies in higher order autoregressive models, where unknown future values of the time series are replaced by forecasts in making forecasts at horizons beyond a single period.

**Forecasting Equity Returns**

This univariate recursive forecasting procedure is easily demonstrated in an empirical example. Consider the log returns to the monthly United States equity index for the period February 1871 to September 2016 expressed in percentage terms as

\[
re_t = 100 \times (\log P_t - \log P_{t-1}),
\]

where \( P_t \) is the equity price index. To generate **ex ante forecasts** of returns using a simple AR(1) model, the parameters are estimated using the entire available sample period and these estimates, together with the actual return for September 2016, are used to generate the recursive forecasts. Consider the case where ex ante forecasts are required for October and November 2016. The estimated model is

\[
re_t = 0.2527 + 0.2839 \, re_{t-1} + \hat{\epsilon}_t,
\]

where \( \hat{\epsilon}_t \) is the least squares fitted residual. Given that the actual return for September 2016 is \(-0.6127\%\), the forecasts for October and November are, respectively,

\[
\begin{align*}
\text{October} & : \hat{r} e_{T+1} = 0.2527 + 0.2839 \, re_T \\
& = 0.2527 + 0.2839 \times -0.6127 = 0.0788\%,
\end{align*}
\]

\[
\begin{align*}
\text{November} & : \hat{r} e_{T+2} = 0.2527 + 0.2839 \, \hat{r} e_{T+1} \\
& = 0.2527 + 0.2839 \times 0.0788 = 0.2751\%.
\end{align*}
\]

Since \( \hat{\phi}_1 = 0.2839 \) and the standard error of the AR(1) regression is \( \hat{\sigma}_e = 3.9036 \), the formulae obtained above for the variances of the optimal forecasts (which ignore the estimation error in the estimate \( \hat{\phi}_1 = 0.2839 \)) may be used to provide a general estimate of the forecast error variances in the above estimates. These are calculated to be 15.2380 for the one period ahead forecast and 16.4663 for the two period ahead forecast. These large forecast error variances indicate that the AR(1) model does not provide particularly informative forecasts for equity returns. Formal ways to evaluate forecasts are discussed in Section 7.5.
7.3 Forecasting Multivariate Time Series Models

The recursive method used to generate the forecasts of a univariate time series model is easily generalised to multivariate models, including VARs and VECMs.

7.3.1 Vector Autoregressions

Consider a bivariate vector autoregression with one lag, VAR(1), given by

\[
\begin{align*}
    y_{1t} &= \phi_{10} + \phi_{11}y_{1t-1} + \phi_{12}y_{2t-1} + v_{1t}, \\
    y_{2t} &= \phi_{20} + \phi_{21}y_{1t-1} + \phi_{22}y_{2t-1} + v_{2t}.
\end{align*}
\]

(7.8)

where \(v_{1t}\) and \(v_{2t}\) are disturbance terms. Given data up to time \(T\), a forecast one period ahead is obtained by writing the model at time \(T+1\)

\[
\begin{align*}
    y_{1T+1} &= \phi_{10} + \phi_{11}y_{1T} + \phi_{12}y_{2T} + v_{1T+1}, \\
    y_{2T+1} &= \phi_{20} + \phi_{21}y_{1T} + \phi_{22}y_{2T} + v_{2T+1}.
\end{align*}
\]

The knowns on the right-hand side are the last observations of the two variables, \(y_{1T}\) and \(y_{2T}\), and the unknowns are the disturbance terms \(v_{1T+1}\) and \(v_{2T+1}\) and the parameters \(\{\phi_{10}, \phi_{11}, \phi_{12}, \phi_{20}, \phi_{21}, \phi_{22}\}\). Replacing the unknowns with estimates, just as in the univariate AR model, the forecasts for the two variables at time \(T+1\) are

\[
\begin{align*}
    \hat{y}_{1T+1} &= \hat{\phi}_{10} + \hat{\phi}_{11}y_{1T} + \hat{\phi}_{12}y_{2T}, \\
    \hat{y}_{2T+1} &= \hat{\phi}_{20} + \hat{\phi}_{21}y_{1T} + \hat{\phi}_{22}y_{2T}.
\end{align*}
\]

To generate forecasts of the VAR(1) model in (7.8) for two periods ahead, the model is written at time \(T+2\)

\[
\begin{align*}
    y_{1T+2} &= \phi_{10} + \phi_{11}y_{1T+1} + \phi_{12}y_{2T+1} + v_{1T+2}, \\
    y_{2T+2} &= \phi_{20} + \phi_{21}y_{1T+1} + \phi_{22}y_{2T+1} + v_{2T+2}.
\end{align*}
\]

All terms on the right-hand side are now unknown. As before the parameters are replaced by estimates and the disturbances are replaced by their means, while \(y_{1T+1}\) and \(y_{2T+1}\) are replaced by their forecasts from the previous step, resulting in the two-period ahead forecasts

\[
\begin{align*}
    \hat{y}_{1T+2} &= \hat{\phi}_{10} + \hat{\phi}_{11}\hat{y}_{1T+1} + \hat{\phi}_{12}\hat{y}_{2T+1}, \\
    \hat{y}_{2T+2} &= \hat{\phi}_{20} + \hat{\phi}_{21}\hat{y}_{1T+1} + \hat{\phi}_{22}\hat{y}_{2T+1}.
\end{align*}
\]

In general, the forecasts of the VAR(1) model for \(H\) periods ahead are

\[
\begin{align*}
    \hat{y}_{1T+H} &= \hat{\phi}_{10} + \hat{\phi}_{11}\hat{y}_{1T+H-1} + \hat{\phi}_{12}\hat{y}_{2T+H-1}, \\
    \hat{y}_{2T+H} &= \hat{\phi}_{20} + \hat{\phi}_{21}\hat{y}_{1T+H-1} + \hat{\phi}_{22}\hat{y}_{2T+H-1}.
\end{align*}
\]
An important new feature of this result is that even if forecasts are required for just one of the variables, say $y_{1t}$, it is necessary to generate forecasts of the other variables in the model as well.

To illustrate forecasting using a VAR consider log returns to equity, $re_t$, and log returns to dividends, $rd_t$, defined in percentage terms as follows

$\begin{align*}
p_t &= 100 \times \log P_t, \\
r_t &= p_t - p_{t-1}, \\
d_t &= 100 \times \log D_t, \\
r_d &= d_t - d_{t-1},
\end{align*}$

where $P_t$ is the equity price index and $D_t$ is the dividend payment. As before data are available for the period February 1871 to September 2016 and suppose ex ante forecasts are required for October and November 2016. The estimated bivariate VAR model is

$\begin{align*}
re_t &= 0.2216 + 0.2836 \cdot re_{t-1} + 0.1058 \cdot rd_{t-1} + \hat{v}_{1t}, \\
r_d &= 0.0316 + 0.0029 \cdot re_{t-1} + 0.8902 \cdot rd_{t-1} + \hat{v}_{2t},
\end{align*}$

where $\hat{v}_{1t}$ and $\hat{v}_{2t}$ are the residuals from the two equations. The forecasts for equity and dividend returns in October are

$\begin{align*}
\hat{re}_{T+1} &= 0.2216 + 0.2836 \cdot re_T + 0.1058 \cdot rd_T + \\
&= 0.2216 + 0.2836 \times -0.6127 + 0.1058 \times 0.4228 \\
&= 0.0926\%,
\end{align*}$

$\begin{align*}
\hat{rd}_{T+1} &= 0.0316 + 0.0029 \cdot re_T + 0.8902 \cdot rd_T + \\
&= 0.0316 + 0.0029 \times -0.6127 + 0.8902 \times 0.4228 \\
&= 0.4062\%.
\end{align*}$

The corresponding forecasts for November are

$\begin{align*}
\hat{re}_{T+2} &= 0.2216 + 0.2836 \cdot \hat{re}_{T+1} + 0.1058 \cdot \hat{rd}_{T+1} \\
&= 0.2216 + 0.2836 \times 0.0926 + 0.1058 \times 0.4062 \\
&= 0.2908\%,
\end{align*}$

$\begin{align*}
\hat{rd}_{T+2} &= 0.0316 + 0.0029 \cdot \hat{re}_{T+1} + 0.8902 \cdot \hat{rd}_{T+1} \\
&= 0.0316 + 0.0029 \times 0.0926 + 0.8902 \times 0.4062 \\
&= 0.3934\%.
\end{align*}$

### 7.3.2 Vector Error Correction Models

An important relationship between systems of VARs and VECMs that was discussed in Chapter 6 is that a VECM may be considered as a restricted VAR
model where the restrictions are embodied in a reduced rank coefficient matrix in the VAR. This correspondence means that a VECM can be re-expressed in VAR form which, in turn, can be used to forecast the variables of the model.

To illustrate, consider the following bivariate VECM which is a restricted version of equation (6.12) in which \( k = 1 \) and \( \delta_1 = \delta_2 = 0 \),

\[
\begin{align*}
\Delta y_{1t} & = a_1 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1}) + \gamma_{11} \Delta y_{1t-1} + \gamma_{12} \Delta y_{2t-1} + \nu_{1t}, \\
\Delta y_{2t} & = a_2 (y_{1t-1} - \beta_0 - \beta_2 y_{2t-1}) + \gamma_{21} \Delta y_{1t-1} + \gamma_{22} \Delta y_{2t-1} + \nu_{2t},
\end{align*}
\]

Rearranging the VECM as a (restricted) VAR(2) in the levels of the variables, gives

\[
\begin{align*}
y_{1t} & = -a_1 \beta_0 + (1 + \gamma_{11} + a_1) y_{1t-1} - \gamma_{11} y_{1t-2} + (\gamma_{12} - a_1 \beta_2) y_{2t-1} - \gamma_{12} y_{2t-2} + \nu_{1t}, \\
y_{2t} & = -a_2 \beta_0 + (\gamma_{21} + a_2) y_{1t-1} - \gamma_{21} y_{1t-2} + (1 + \gamma_{22} - a_2 \beta_2) y_{2t-1} - \gamma_{22} y_{2t-2} + \nu_{2t}.
\end{align*}
\]

Alternatively, it is possible to write this system in VAR form as

\[
\begin{align*}
y_{1t} & = \phi_{10} + \phi_{11} y_{1t-1} + \phi_{12} y_{1t-2} + \phi_{13} y_{2t-1} + \phi_{14} y_{2t-2} + \nu_{1t}, \\
y_{2t} & = \phi_{20} + \phi_{21} y_{1t-1} + \phi_{22} y_{1t-2} + \phi_{23} y_{2t-1} + \phi_{24} y_{2t-2} + \nu_{2t},
\end{align*}
\]

(7.9)

in which the VAR and VECM parameters are related as follows

\[
\begin{align*}
\phi_{10} & = -a_1 \beta_0 & \phi_{20} & = -a_2 \beta_0 \\
\phi_{11} & = 1 + a_1 + \gamma_{11} & \phi_{21} & = -a_2 + \gamma_{21} \\
\phi_{12} & = -\gamma_{11} & \phi_{22} & = -\gamma_{21} \\
\phi_{13} & = -a_1 \beta_2 + \gamma_{12} & \phi_{23} & = 1 - a_2 \beta_2 + \gamma_{22} \\
\phi_{14} & = -\gamma_{12} & \phi_{24} & = -\gamma_{22}.
\end{align*}
\]

(7.10)

These equations give the explicit form of the reduced rank restrictions implicit in the VECM system. Once the VECM is expressed as a VAR in the levels of the variables as in equation (7.9), forecasts are generated for the VAR as in Section 7.3.1 with the VAR parameter estimates computed from the VECM parameter estimates based on the explicit relationships given in (7.10).

Using the same dataset as that used in producing the ex ante VAR forecasts in the previous section, the procedure is easily repeated for the VECM. The estimated VECM with one lag and also relaxing the restriction \( \delta_1 = \delta_2 = 0 \) is

\[
\begin{align*}
re_t & = 0.0353 - 0.0068(p_{t-1} - 1.1773 d_{t-1} - 340.3414) \\
& + 0.2901 re_{t-1} + 0.1366 rd_{t-1} + \tilde{v}_{1t}, \\
r d_{t} & = 0.0984 + 0.0024(p_{t-1} - 1.1773 d_{t-1} - 340.3414) \\
& + 0.0006 re_{t-1} + 0.8792 rd_{t-1} + \tilde{v}_{2t},
\end{align*}
\]

(7.11)

\[\text{These estimates are the same as the estimates reported in Chapter 6 with the exception that the intercepts now reflect the fact that the variables are scaled by 100.}\]
where \( \hat{v}_1 \) and \( \hat{v}_2 \) are the residuals from the two equations. Since \( r_e_t = p_t - p_{t-1} \) and \( r_d_t = d_t - d_{t-1} \), the VECM(1) is rewritten as a VAR(2) in levels as follows

\[
\begin{align*}
    p_t &= (0.0353 + 0.0068 \times 340.3414) \\
         &+ (1 - 0.0068 + 0.2901) p_{t-1} - 0.2901 p_{t-2} \\
         &+ (0.0068 \times 1.1773 + 0.1366) d_{t-1} - 0.1366 d_{t-2} + \hat{v}_{1t}, \\
    d_t &= (0.0984 + 0.0024 \times 340.3414) \\
         &+ (0.0024 + 0.0006) p_{t-1} - 0.0006 p_{t-2} \\
         &+ (1 - 0.0024 \times 1.1773 + 0.8792) d_{t-1} - 0.8792 d_{t-2} + \hat{v}_{2t},
\end{align*}
\]

or

\[
\begin{align*}
    p_t &= 2.3350 + 1.2980 p_{t-1} - 0.2901 p_{t-2} + 0.1433 d_{t-1} - 0.1366 d_{t-2} + \hat{v}_{1t}, \\
    d_t &= 0.9232 + 0.0035 p_{t-1} - 0.0006 p_{t-2} + 1.8816 d_{t-1} - 0.8792 d_{t-2} + \hat{v}_{2t}.
\end{align*}
\]

The forecast for the October 2016 log equity price and log dividend payment given their values in August, \( p_{T-1} \) and \( d_{T-1} \), and September, \( p_T \) and \( d_T \), are obtained using these equations as

\[
\begin{align*}
    \hat{p}_{T+1} &= 2.3350 + 1.2980 p_T - 0.2901 p_{T-1} + 0.1433 d_T - 0.1366 d_{T-1} \\
                 &= 767.7360,
\end{align*}
\]

and

\[
\begin{align*}
    \hat{d}_{T+1} &= 0.9232 + 0.0035 p_T - 0.0006 p_{T-1} + 1.8816 d_T - 0.8792 d_{T-1} \\
                 &= 381.1519.
\end{align*}
\]

Similar calculations reveal that the respective forecasts for the November 2016 log equity price and log dividend payment (both scaled by 100) are

\[
\begin{align*}
    \hat{p}_{T+2} &= 767.9893, \\
    \hat{d}_{T+2} &= 381.5670.
\end{align*}
\]

Based on these forecasts, percentage equity returns in October and November are, respectively,

\[
\begin{align*}
    \hat{r}_{eT+1} &= 767.7360 - 767.6793 = 0.0567\%, \\
    \hat{r}_{eT+2} &= 767.9893 - 767.7360 = 0.2533\%,
\end{align*}
\]

and the corresponding forecasts for dividend returns are, respectively,

\[
\begin{align*}
    \hat{r}_{dT+1} &= 381.1519 - 380.7329 = 0.4190\%, \\
    \hat{r}_{dT+2} &= 381.5670 - 381.1519 = 0.4151\%.
\end{align*}
\]
7.4 Combining Forecasts

Models are stylised representations of real world phenomena that, at best, capture only approximately the complexities of the underlying processes. Even though models can never be expected to represent the true generating process exactly, some models can be very useful in modelling and forecasting economic and financial variables. In approximating the true process and in forecasting future realisations, certain models may prove to be better than others. To take advantage of the good properties of certain models and the robustness of others, it is sometimes helpful to combine models and the forecasts from these models in a manner that exploits their individual properties.

Such combinations have often been found in practical work to reduce forecast error variance. It is not surprising, therefore, that the concept and methodology of forecast combination has attracted much academic interest. Timmerman (2006) and Elliott and Timmerman (2008, 2016) provide extensive references to the vast literature on this subject. In practical forecasting work, the media often report consensus forecasts of key economic and financial variables such as GDP and inflation that are important in financial decision making. Such forecasts are obtained by taking simple averages of forecasts of the same quantity that are provided by different forecasting units or different forecasting techniques. The availability of multiple forecasts raises an important question in forecasting: is it better to rely on an individual forecast with known good properties or are there potential gains to averaging several competing forecasts?

Suppose that two unbiased forecasts of a variable $y_t$ are available, given by $\hat{y}^1_t$ and $\hat{y}^2_t$, with respective forecast variances $\sigma^2_1$ and $\sigma^2_2$ and covariance $\sigma_{12}$. A weighted average of these two forecasts is

$$\hat{y}_t = \omega \hat{y}^1_t + (1 - \omega) \hat{y}^2_t,$$

which depends on a weight parameter $\omega \in [0, 1]$. The combined forecast $\hat{y}_t$ is also unbiased and its variance is

$$\sigma^2 = \omega^2 \sigma^2_1 + (1 - \omega)^2 \sigma^2_2 + 2\omega(1 - \omega)\sigma_{12}.
$$

A natural approach to selecting the combined forecast is to choose the weight $\omega$ in order to minimise its forecast variance. The first-order condition for a minimum is given by

$$\frac{d\sigma^2}{d\omega} = 2\omega \sigma^2_1 - 2(1 - \omega) \sigma^2_2 + 2\sigma_{12} - 4\omega \sigma_{12},$$

and setting this expression to zero and solving gives the optimal weight parameter

$$\omega = \frac{\sigma^2_2 - \sigma_{12}}{\sigma^2_1 + \sigma^2_2 - 2\sigma_{12}}.$$
It is clear therefore that the weight attached to \( \hat{y}_t \) varies inversely with its variance. In passing, it is worth noting that these weights are identical to the optimal weights for the minimum variance portfolio derived in Chapter 3.

This point can be illustrated more clearly if the forecasts are assumed to be uncorrelated so that \( \sigma_{12} = 0 \). In this case,

\[
\omega = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}, \quad 1 - \omega = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2},
\]

and it is clear that both forecasts have weights varying inversely with their individual variances. By rearranging the expression for \( \omega \) as follows

\[
\omega = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} \left( \frac{\sigma_2^{-2} \sigma_1^{-2}}{\sigma_2^{-2} \sigma_1^{-2} + \sigma_2^{-2}} \right) = \frac{\sigma_1^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}}, \tag{7.11}
\]

the inverse proportionality is now manifest in the numerator of expression (7.11). The inverse of the forecast variance, \( \sigma_i^{-2} \), may be interpreted as a measure of the imprecision of the forecast \( \hat{y}_t \) (that is, the larger the variance, the lower the precision of the forecast). The implication of (7.11) is that the lower the precision of the forecast \( \hat{y}_t \) (relative to the overall precision) the less weight (\( \omega \)) is placed on that forecast in the combination.

This simple intuition in the two forecast case translates into a situation in which there are \( N \) forecasts \( \{ \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_N \} \) of the same variable \( y_t \). If these forecasts are all unbiased and uncorrelated and if the weights satisfy

\[
\sum_{i=1}^{N} \omega_i = 1 \quad \omega_i \geq 0 \quad i = 1, 2, \ldots, N,
\]

then from (7.11) the optimal weights are

\[
\omega_i = \frac{\sigma_i^{-2}}{\sum_{j=1}^{N} \sigma_j^{-2}},
\]

and the weight on forecast \( i \) is inversely proportional to its variance.

The weights in expression (7.11) are intuitively appealing as they are based on the principle of producing a minimum variance portfolio of the forecasts. Important questions remain, however, about how best to implement the combination of forecasts approach in practice. Bates and Granger (1969) suggested using (7.11) to construct the weights with the required estimates of the forecast variances, \( \hat{\sigma}_i^2 \), given by the forecast mean square error based, for instance, on within sample performance. All this approach requires then is an estimate of the mean square error of all the competing forecasts in order to compute estimates of the optimal weights, \( \hat{\omega}_i \). Granger and Ramanathan (1984) later show that this method is numerically equivalent to weights constructed from running the restricted regression

\[
y_t = \omega_1 \hat{y}_1^t + \omega_2 \hat{y}_2^t + \cdots + \omega_N \hat{y}_N^t + v_t,
\]
in which the constant term is zero and the coefficients are constrained to be non-negative and to sum to one. Enforcing these restrictions involves constrained optimisation that can be difficult in practice and sometimes ad hoc methods are adopted instead. One method is the sequential elimination of forecasts with weights estimated to be negative until all the remaining forecasts in the proposed combination forecast have positive weights.

Yet another approach to averaging forecasts is based on the use of information criteria (Buckland, Burnham and Augustin, 1997; Burnham and Anderson, 2002), which may be interpreted as an estimate of the relative quality of an econometric model. Suppose there are $N$ different models each with an estimated Akaike information criterion $AIC_1, AIC_2, \cdots, AIC_N$, then the model that returns the minimum value of the information criterion is usually the model of choice. Denote the minimum value of the information criterion for this set of models as $AIC_{\text{min}}$. Then the expression

$$\exp \left[ \Delta I_i / 2 \right] = \exp \left[ (AIC_i - AIC_{\text{min}}) / 2 \right],$$

may be interpreted as a relative measure of the loss of information from using model $i$ instead of the model that produces $AIC_{\text{min}}$. It is therefore natural to allow the forecast combination to reflect this relative information by computing the weights in terms of the ratios

$$\hat{\omega}_i = \frac{\exp \left[ \Delta I_i / 2 \right]}{\sum_{j=1}^{N} \exp \left[ \Delta I_j / 2 \right]}.$$

The Schwarz (Bayesian) Information Criterion (SIC) has also been suggested as an alternative information criterion to use in the application of this method (Garratt, Koop and Vahey, 2008; Kapetanios, Labhard and Price, 2008).

The simplest approach of all is to assign equal weights to these forecasts and construct the simple average

$$\hat{y}_t = \frac{1}{N} \sum_{i=1}^{N} \hat{y}_{1it}.$$

Interestingly, simulation studies and practical work indicate that this simplistic strategy of averaging forecasts often works better than other more sophisticated methods, especially when there are large numbers of forecasts to be combined, notwithstanding all the subsequent work on the optimal estimation of weights (Stock and Watson, 2001). Two possible explanations for why simple averaging might work better in practice than constructing an optimal combination with estimated weights are as follows.

---

2 The exact form of this expression derives from the likelihood principle which is discussed in Chapter 10. The AIC is an unbiased estimator of $-2$ times the log-likelihood function of model $i$, so that after dividing by $-2$ and exponentiating, the result is a measure of the likelihood that model $i$ actually generated the observed data.
(i) There may be significant errors in the estimation of the weights, due to either parameter instability (Clemen, 1989; Winkler and Clemen, 1992, Smith and Wallis, 2009) or structural breaks (Hendry and Clements, 2004).

(ii) If the forecast variances of the competing forecasts are similar and their covariances are positive, then large gains obtained by constructing optimal weights are unlikely (Elliott, 2011). Note, for example, in the two forecast combination case, if the forecast variances are the same, the weight formula is

$$\omega = \frac{\sigma^2 - \sigma_{12}}{2\sigma^2 - 2\sigma_{12}} = \frac{1}{2},$$

delivering equal weights.

### 7.5 Forecast Evaluation Statistics

The discussion so far has concentrated on ex ante forecasting of a variable or variables over a forecast horizon, $H$, beginning after the last observation in the dataset. However, it is also of interest to be able to compare the forecasts with the actual values that are observed in order to assess accuracy. One approach is simply to wait until the future values are observed, but this is not convenient if information about the forecasting capability of a model is required before ex ante forecasting is conducted.

A common solution that is adopted to determine the forecast accuracy of a model is to estimate the model over a restricted sample period that excludes the last $H$ observations. The model is then forecasted out-of-sample over these observations. Since the actual values of the variables over these $H$ time periods have already been observed it is now possible to compare forecasts with historical values to assess accuracy. Forecasts computed in this way are known as ex post forecasts according to the definition given earlier.

Suppose now that ex post forecasts are required for the period January 2016 to September 2016 for United States equity returns, $re_t$, using the same data set used to generate the ex ante forecasts of the previous section based on the AR(1) model. The model is estimated over the restricted period February 1871 to December 2015 to yield

$$re_t = 0.2518 + 0.2840 re_{t-1} + \hat{\epsilon}_t,$$

where $\hat{\epsilon}_t$ is the least squares residual. The forecasts are now generated recursively using the estimated model and also the fact that the equity return in December 2015 is $re_{T-H} = -1.2838\%$:

Jan:  \( \hat{r}e_{T-H+1} = 0.2518 + 0.2840 \hat{r}e_{T-H} \)
7.5. **FORECAST EVALUATION STATISTICS**

\[ \hat{r}_{T-H+2} = 0.2518 + 0.2840 \times -1.2838 = -0.1127\% , \]

Feb: \[ \hat{r}_{T-H+3} = 0.2518 + 0.2840 \times -0.1127 = 0.2198\% , \]

Mar: \[ \hat{r}_{T-H+4} = 0.2518 + 0.2840 \times 0.2198 = 0.3141\% , \]

Apr: \[ \hat{r}_{T-H+5} = 0.2518 + 0.2840 \times 0.3143 = 0.3411\% , \]

May: \[ \hat{r}_{T-H+6} = 0.2518 + 0.2840 \times 0.3411 = 0.3487\% , \]

Jun: \[ \hat{r}_{T-H+7} = 0.2518 + 0.2840 \times 0.3487 = 0.3509\% , \]

Jul: \[ \hat{r}_{T-H+8} = 0.2518 + 0.2840 \times 0.3509 = 0.3515\% , \]

Aug: \[ \hat{r}_{T-H+9} = 0.2518 + 0.2840 \times 0.3517 = 0.3517\% . \]

The ex post forecasts of United States equity returns are illustrated in Figure 7.1. It is readily apparent how quickly the forecasts are driven toward the unconditional mean of returns 0.3539%. As pointed out previously, this pattern is typical of time series forecasts based on stationary data.

There are a number of simple summary statistics that are used to determine the accuracy of ex post forecasts. Define the forecast errors as the differences between the actual values and the forecasted values over the forecast horizon, namely

\[ y_{T+1} - \hat{y}_{T+1}, y_{T+2} - \hat{y}_{T+2}, \ldots, y_{T+H} - \hat{y}_{T+H}. \]

The smaller are these forecast errors, the better are the forecasts. So the differences may be used to compute a summary statistic. The most commonly used summary measures of overall closeness of the forecasts to the actual values are the mean absolute error (MAE), mean absolute percentage error (MAPE), the mean square error (MSE) and the root mean square error (RMSE). These metrics are defined as follows:

\[ MAE = \frac{1}{H} \sum_{h=1}^{H} |y_{T+h} - \hat{y}_{T+h}|, \]

\[ MAPE = \frac{1}{H} \sum_{h=1}^{H} \left| \frac{y_{T+h} - \hat{y}_{T+h}}{y_{T+h}} \right|, \]

\[ MSE = \frac{1}{H} \sum_{h=1}^{H} (y_{T+h} - \hat{y}_{T+h})^2, \]
Figure 7.1: Ex post forecasts (shown by the dashed line) of United States equity returns (%) generated by an AR(1) model. The estimation sample period is February 1871 to December 2015 and the forecast period is from January 2016 to September 2016.

\[
RMSE = \sqrt{\frac{1}{H} \sum_{h=1}^{H} (y_{T+h} - \hat{y}_{T+h})^2}.
\]

To compute the MSE for the forecast period the actual sample observations of equity returns from January 2016 to September 2016 are required. These percentage returns are

\[1.8746, 1.0696, 2.8859, 4.6891, 0.9526, -1.7095, 0.8311, -2.7352, 2.6822.\]

The MSE is

\[
MSE = \frac{1}{9} \sum_{h=1}^{9} (r_{T+h} - \hat{r}_{T+h})^2
\]

\[
= \frac{1}{9} \left( (1.8746 - (-0.1127))^2 + (1.0696 - 0.2198)^2 + (2.8859 - 0.3143)^2 \\
+ (4.6891 - 0.3411)^2 + (0.9526 - 0.3487)^2 + (-1.7095 - 0.3509)^2 \\
+ (0.8311 - 0.3515)^2 + (-2.7352 - 0.3517)^2 + (2.6822 - 0.3517)^2 \right)
\]

\[
= 10.3423.
\]
The RMSE is
\[
RMSE = \sqrt{\frac{1}{9} \sum_{h=1}^{9} (r_{t+h} - \hat{r}_{t+h})^2} = \sqrt{10.3423} = 3.2159.
\]

Taken on its own, the root mean squared error of the forecast, 3.2159, does not provide a good descriptive measure of the relative accuracy of this model because its value can be changed by simply changing the units of measurement of the data. For example, expressing the data as returns and not as percentage returns results in the RMSE falling by a factor of 100. Clearly this smaller RMSE does not mean that the forecasting performance of the AR(1) model has improved. The way that the RMSE and the MSE are typically used to evaluate the forecasting performance of a model is to compute the same statistics for an alternative model. The model with the smaller RMSE or MSE is judged to be the better forecasting model.

The forecasting performance characteristics of several models are now compared in a practical exercise. The models employed are an AR(1) model of equity returns, a VAR(1) model containing both equity and dividend returns, and a VECM(2) with an unrestricted constant, containing log equity prices and log dividend payments. Each model is estimated using a reduced sample on United States monthly percentage equity returns from February 1871 to December 2015 and used to generate forecasts from January to September 2016. The resulting forecasts are then compared using the MSE and RMSE statistics.

Table 7.1
Forecasting performance of models of United States monthly percentage equity returns. All models are estimated over the period January 1871 to December 2015 and forecasts are from January to September 2016.

<table>
<thead>
<tr>
<th></th>
<th>2016</th>
<th>2016</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AR(1)</td>
<td>VAR(1)</td>
<td>VECM(2)</td>
</tr>
<tr>
<td>Jan</td>
<td>-0.1127%</td>
<td>-0.0713%</td>
<td>-0.1013%</td>
</tr>
<tr>
<td>Feb</td>
<td>0.2198%</td>
<td>0.2678%</td>
<td>0.2426%</td>
</tr>
<tr>
<td>Mar</td>
<td>0.3143%</td>
<td>0.3599%</td>
<td>0.3410%</td>
</tr>
<tr>
<td>Apr</td>
<td>0.3411%</td>
<td>0.3825%</td>
<td>0.3677%</td>
</tr>
<tr>
<td>May</td>
<td>0.3487%</td>
<td>0.3859%</td>
<td>0.3737%</td>
</tr>
<tr>
<td>Jun</td>
<td>0.3509%</td>
<td>0.3841%</td>
<td>0.3738%</td>
</tr>
<tr>
<td>Jul</td>
<td>0.3515%</td>
<td>0.3811%</td>
<td>0.3723%</td>
</tr>
<tr>
<td>Aug</td>
<td>0.3517%</td>
<td>0.3781%</td>
<td>0.3705%</td>
</tr>
<tr>
<td>Sep</td>
<td>0.3517%</td>
<td>0.3753%</td>
<td>0.3687%</td>
</tr>
<tr>
<td>MSE</td>
<td>10.3423</td>
<td>10.3234</td>
<td>10.3076</td>
</tr>
<tr>
<td>RMSE</td>
<td>3.2159</td>
<td>3.2130</td>
<td>3.2106</td>
</tr>
</tbody>
</table>

The results in Table 7.1 show that the VECM(2) is the best forecasting model.
as it yields the smallest MSE and RMSE. This result is not surprising given
the earlier discussion in Chapter 6 which revealed the usefulness of a cointe-
grating link in capturing the interdependence between dividends and equity
prices. The VECM is designed to capture this long-term linkage and use the
link in forecasting. The fact that the VAR(1), which is formulated with both
equity returns and dividend returns data, provides the second best forecasting
performance suggests that transient linkages between these variables is
also important. In particular, the findings indicate that accounting for short-
run dynamic feedback effects between equity returns and dividend returns is
worth doing even when long-term linkages are ignored, giving forecast per-
formance that is superior to a univariate specification for equity returns.

A widely used formal test for comparing two different forecasts is due to
Diebold and Mariano (1995). Suppose we have two competing forecasts and
we are able to compute the forecast error, $\hat{y}_{T+H}(M_j) = y_{T+H} - \hat{y}_{T+H}$, for the
$j$th model, $M_j$. Now define the difference

$$w_t = \hat{u}_{T+H}(M_1) - \hat{u}_{T+H}(M_2).$$

The Diebold-Mariano test of equal predictive accuracy is based on a simple $t$
test that $E(w_t) = 0$. Proper construction of this $t$ test relies on the use of an
appropriate standard error for the difference $w_t$.

There is an active research area in financial econometrics in which these sta-
tistical (or direct) measures of forecast performance are replaced by problem-
specific (or indirect) measures of forecast performance in which the evaluation
relates specifically to some relevant economic decision (Elliott and Timmer-
man, 2008; Patton and Sheppard, 2009). Examples of the indirect approach to
forecast evaluation appear in Engle and Colacito (2006), who evaluate forecast
performance in terms of portfolio return variance, and Fleming, Kirby and
Ostdiek (2001, 2003), who apply a quadratic utility function that values one
forecast relative to another. Becker, Clements, Doolan and Hurn (2013) survey
and compare these different approaches to forecast evaluation.

### 7.6 Evaluating the Density of Forecast Errors

The above discussion of forecast generation for financial variables has fo-
cused on first and second moment properties: the conditional mean point
forecast and the conditional variance of the forecast distribution, which can
both be used to help construct an interval forecast. A natural extension is to
include forecasts of higher order moments such as skewness and kurtosis or
even the entire probability distribution. The latter is of particular interest in
the area of risk management where assessments of such quantities as future
Value at Risk are relevant.

As is the case with point forecasts, where statistics are computed to determine
the relative accuracy of the forecasts, density forecasts may also be evaluated
7.6. EVALUATING THE DENSITY OF FORECAST ERRORS

to determine their relative accuracy. The approach that is typically used is to consider the entire distribution by using a quantity known as the probability integral transform (PIT), which is now discussed.

7.6.1 Probability Integral Transform

To fix ideas in a simple case, consider the constant mean model of returns given by

\[ y_t = \mu + v_t, \quad v_t \sim N(0, 1), \quad (7.12) \]

in which \( \mu = 0 \). Denote the cumulative distribution function (cdf) of the standard normal distribution evaluated at any point \( z \) as \( \Phi(z) \). If the observed values \( y_t \) are indeed generated correctly according to this simple model, then the transformed quantity

\[ u_t = \Phi \left( \frac{y_t - \mu}{\sigma} \right), \quad t = 1, 2, \cdots, T, \quad (7.13) \]

takes values in the unit interval \([0, 1]\) because of the fundamental property of the cdf, \( \Phi \). Furthermore, and again assuming that \( y_t \) is generated by (7.12), the transformed time series \( u_t \) has a standard uniform distribution on this interval because

\[ P(u_t \leq u) = P \left( \Phi \left( \frac{y_t - \mu}{\sigma} \right) \leq u \right) \]
\[ = P \left( \frac{y_t - \mu}{\sigma} \leq \Phi^{-1}(u) \right) \]
\[ = \Phi(\Phi^{-1}(u)) = u. \quad (7.14) \]

The transformation (7.13) is known as the probability integral transform because use of the cdf \( \Phi \) in (7.13) involves a transform of probability as shown in (7.14).

Figure 7.2 illustrates how the transformed times series \( u_t \) is obtained from the actual time series \( y_t \) where the specified model is \( N(0, 1) \). This result reflects the property that if the cumulative distribution is indeed the correct distribution, transforming \( y_t \) to \( u_t \) means that each transformed quantity \( u_t \) has the same probability of being realised as any other value of \( u_t \), as corroborated directly by (7.14).

The probability integral transform in the case where the specified model is chosen correctly is highlighted in panel (a) of Figure 7.3. A time series plot of 1000 simulated observations, \( y_t \), drawn from a \( N(0, 1) \) distribution is transformed into \( u_t \) using the cumulative normal distribution in (7.13). In the third column of panel (a) the histogram of the transformed time series \( u_t \) is shown. Inspection of the histogram corroborates empirically that the distribution used in transforming \( y_t \) is indeed the correct one because distribution of \( u_t \) is close to uniform over the whole interval \([0, 1]\).
Next consider the case where the true data generating process for \( y_t \) is a \( N(0.5, 1) \) distribution, but the incorrect distribution based on (7.12) is used as the forecast distribution to perform the probability integral transform. The effect of misspecifying the mean of the forecast distribution is illustrated in panel (b) of Figure 7.3. A time series of 1000 simulated observations from a \( N(0.5, 1.0) \) distribution, \( y_t \), is transformed using the incorrect distribution, \( N(0, 1) \), and the histogram of the transformed time series, \( u_t \), is plotted in the third column. The histogram of \( u_t \) appears to rise systematically over the interval \([0, 1]\) and clearly departs from a uniform distribution. The departure from uniformity in this case is a reflection of the misspecified mean model in the forecasting model. The histogram exhibits a positive slope reflecting the fact that larger values of \( u_t \), and hence \( y_t \), have relatively higher probability of occurrence than smaller values of \( u_t \) and \( y_t \).

Finally consider the case where the variance of the model is misspecified. If the data generating process is a \( N(0, 2) \) distribution, but the forecast distribution used in the probability integral transform is again \( N(0, 1) \), then it is to be expected that the forecast distribution will understate the true spread of the data. This is visible in panel (c) of Figure 7.3. The histogram of \( u_t \) is now U-shaped implying that large negative and large positive values have a higher probability of occurring than are predicted by the \( N(0,1) \) distribution.
7.6. EVALUATING THE DENSITY OF FORECAST ERRORS

Figure 7.3: Simulated time series to show the effects of misspecification using the probability integral transform. In panel (a) there is no misspecification while panels (b) and (c) demonstrate the effect of misspecification in the mean and variance of the distribution respectively.

7.6.2 Equity Returns

The models used to forecast United States equity log returns \( r_t \) in Section 7.2 are all based on the assumption of normality. Consider the AR(1) model

\[
r_t = \phi_0 + \phi_1 r_{t-1} + \nu_t, \quad \nu_t \sim N(0, \sigma^2_v).
\]

Assuming the forecast is ex post so that \( r_t \) is available, the one-step ahead forecast error is given by

\[
\hat{\nu}_t = r_t - \hat{\phi}_0 - \hat{\phi}_1 r_{t-1},
\]

with asymptotic distribution

\[
\hat{\nu}_t \sim N(r_t - \phi_0 - \phi_1 r_{t-1}, \sigma^2_v),
\]

(7.15)
where the parameters are estimated using monthly data from January 1871 to September 2016. For the purpose of this exercise, the calculations do not take into account the distributional effects of the estimation error in the fitted coefficients ($\hat{\phi}_0, \hat{\phi}_1$). The distribution in (7.15) is

$$\tilde{v}_t \sim N(r_t - 0.2527 - 0.2839 r_{t-1}, 3.9036^2).$$

The probability integral transform corresponding to the estimated distribution in (7.15) is then

$$u_t = \Phi\left(\frac{\tilde{v}_t}{\hat{\sigma}_v}\right),$$

in which $\hat{\sigma}_v = 3.9036$ is the standard error of the regression. A histogram of the transformed time series, $u_t$, is given in Figure 7.4. It appears that the AR(1) forecasting model of equity returns is misspecified because the distribution of $u_t$ is non-uniform. The interior peak of the distribution of $u_t$ suggests that the distribution of $y_t$ is more peaked than that predicted by the normal distribution. Also, the peak in the distribution at zero suggests that there are some observed large negative values of $y_t$ that are not consistent with the specification of a normal distribution. These two properties together indicate that the specified model fails to take into account the presence of skewness and kurtosis in the actual data.

The analysis of the one-step ahead AR(1) forecasting model can easily be extended to the other estimated models of equity returns including the VAR model and the VECM investigated in Section 7.3 to forecast equity returns. As applied here the probability integral transform is performed ex post because it uses within sample one-step ahead prediction errors to conduct the analysis. The application is also a graphical implementation in which misspecification is detected by simple visual inspection of the histogram of the transformed time series $u_t$. It is possible to relax both these limitations in practice. In particular, Diebold, Gunther and Tsay (1998) discuss an alternative approach suited to ex ante applications, and Ghosh and Bera (2005) propose a class of formal statistical tests of the null hypothesis that $u_t$ is uniformly distributed.

### 7.7 Regression Model Forecasts

The forecasting techniques for univariate and multivariate models that have been discussed so far are all based on time series models in which each dependent variable is expressed as a function of own lags and lags of other variables. Additional explanatory variables may be used in forecasting exercises. A simple framework for using explanatory variables in forecasting exercises is the linear regression model

$$y_t = \beta_0 + \beta_1 x_t + u_t,$$
where $y_t$ is the dependent variable, $x_t$ is the explanatory variable, $u_t$ is a disturbance term, and the sample period is $t = 1, 2, \cdots, T$. To generate a forecast of $y_t$ at time $T + 1$, as before, the model is written at $T + 1$ as

$$y_{T+1} = \beta_0 + \beta_1 x_{T+1} + u_{T+1}. \tag{7.16}$$

The unknown values on the right hand-side are $x_{T+1}$ and $u_{T+1}$, as well as the parameters $\{\beta_0, \beta_1\}$. As before, the equation error in the forecast period $u_{T+1}$ is replaced by its expected value of $E[u_{T+1}] = 0$, and the parameters are replaced by their sample estimates, $(\hat{\beta}_0, \hat{\beta}_1)$. However, unlike autoregressive formulations in which lagged values of the dependent variable are observed in a one-period ahead forecast, the future value of the explanatory variable $x_{T+1}$ is unobserved.

One strategy is to specify hypothetical future values of the explanatory variable that represent potential outcomes of interest. A less subjective approach is to specify a time series model for $x_t$ and use this model to generate forecasts of future values $x_{T+H}$ by means of the methods discussed previously. For example, if an AR(2) model is proposed for $x_t$ the overall model becomes a bivariate system of equations of the form

$$y_t = \beta_0 + \beta_1 x_t + u_t, \tag{7.16}$$

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + v_t. \tag{7.17}$$
CHAPTER 7. FORECASTING

To generate the first forecast at time $T + 1$ the system of equations is written as

\[
\begin{align*}
y_{T+1} &= \beta_0 + \beta_1 x_{T+1} + u_{T+1}, \\
x_{T+1} &= \phi_0 + \phi_1 x_T + \phi_2 x_{T-1} + v_{T+1}.
\end{align*}
\]

Replacing the unknown parameters with estimates and the unknown equation errors with expectations yields the predictive system

\[
\begin{align*}
\hat{y}_{T+1} &= \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_{T+1}, \\
\hat{x}_{T+1} &= \hat{\phi}_0 + \hat{\phi}_1 x_T + \hat{\phi}_2 x_{T-1}.
\end{align*}
\]

Equation (7.19) is used to generate the forecast $\hat{x}_{T+1}$, which is utilised in equation (7.18) to generate $\hat{y}_{T+1}$.

These calculations can be performed in a single step by substituting (7.19) for $\hat{x}_{T+1}$ into (7.18) to give

\[
\hat{y}_{T+1} = \hat{\beta}_0 + \hat{\beta}_1 (\hat{\phi}_0 + \hat{\phi}_1 x_T + \hat{\phi}_2 x_{T-1}) = \hat{\beta}_0 + \hat{\beta}_1 \hat{\phi}_0 + \hat{\beta}_1 \hat{\phi}_1 x_T + \hat{\beta}_1 \hat{\phi}_2 x_{T-1}.
\]

Using the same approach, multiple explanatory variables are easily handled by specifying a VAR to generate the required multivariate forecasts of these variables.

As an illustration, a regression model may be used to forecast United States equity returns, $re_t$, using dividend returns, $rd_t$. As in earlier examples, the data employed are from February 1871 to September 2016. Estimation of equations (7.16) and (7.17) is conducted by ordinary least squares. For simplicity, the model for the explanatory variable is restricted to an AR(1). The fitted system is

\[
\begin{align*}
re_t &= 0.3433 + 0.0360 rd_t + \hat{\nu}_t, \\
rd_t &= 0.0326 + 0.8903 rd_{t-1} + \hat{\nu}_t.
\end{align*}
\]

The forecasts for dividend returns in October and November are then

\[
\begin{align*}
\hat{rd}_{T+1} &= 0.0326 + 0.8903 \hat{rd}_T = 0.0326 + 0.8903 \times 0.4228 = 0.4090\%, \\
\hat{rd}_{T+2} &= 0.0326 + 0.8903 \hat{rd}_{T+1} = 0.0326 + 0.8903 \times 0.4090 = 0.3967\%.
\end{align*}
\]

and the corresponding equity return forecasts are

\[
\begin{align*}
\hat{re}_{T+1} &= 0.3433 + 0.0360 \hat{rd}_{T+1} = 0.3433 + 0.0360 \times 0.3982 = 0.3580\%, \\
\hat{re}_{T+2} &= 0.3433 + 0.0360 \hat{rd}_{T+2} = 0.3433 + 0.0360 \times 0.3871 = 0.3576%.
\end{align*}
\]
7.8 Predicting the Equity Premium

Forecasting in finance using time series regression models is a highly active field of research in applied finance. One area where the methods have been extensively used is in predicting the equity premium. Two influential papers on this topic are by Goyal and Welch (2003, 2008). They address the problem of predicting the equity premium, which is defined as the excess return from investment in equities over the return from a risk free investment. The relevant equations are

\[ EQP_t = rm_t - rf_t \]
\[ rm_t = \log(P_t + D_t) - \log(P_{t-1}) \]
\[ rf_t = \log(1 + Rf_t), \]

where \( P_t \) is an equity price index, \( D_t \) is the dividend payment stream associated with the index, and \( Rf_t \) is a representative risk free interest rate. Two explanatory predictors are used in the regressions, namely the dividend-price ratio and the dividend-yield ratio, which are defined as

\[ DP_t = \log(D_t) - \log(P_t) \]
\[ DY_t = \log(D_t) - \log(P_{t-1}). \]

Table 7.2 provides summary statistics of the annual data used by Goyal and Welch (2003). For these calculations \( P_t \) is the value-weighted CRSP index, \( D_t \) is the dividend paid on the index, \( DP_t \) is the corresponding dividend price ratio, and the equity premium \( EQP_t \) is obtained using the three-month Treasury

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1926 - 2002</td>
<td>rm_t</td>
<td>9.29</td>
<td>19.80</td>
<td>-58.74</td>
<td>45.71</td>
<td>-0.89</td>
</tr>
<tr>
<td></td>
<td>EQP_t</td>
<td>5.57</td>
<td>20.00</td>
<td>-59.82</td>
<td>45.41</td>
<td>-0.75</td>
</tr>
<tr>
<td></td>
<td>DP_t</td>
<td>-3.30</td>
<td>0.42</td>
<td>-4.48</td>
<td>-2.36</td>
<td>-0.68</td>
</tr>
<tr>
<td></td>
<td>DY_t</td>
<td>-3.25</td>
<td>0.40</td>
<td>-4.53</td>
<td>-2.56</td>
<td>-1.17</td>
</tr>
<tr>
<td>1946 - 2002</td>
<td>rm_t</td>
<td>10.34</td>
<td>16.00</td>
<td>-32.72</td>
<td>40.72</td>
<td>-0.52</td>
</tr>
<tr>
<td></td>
<td>EQP_t</td>
<td>5.69</td>
<td>16.37</td>
<td>-40.46</td>
<td>39.86</td>
<td>-0.53</td>
</tr>
<tr>
<td></td>
<td>DP_t</td>
<td>-3.41</td>
<td>0.41</td>
<td>-4.48</td>
<td>-2.73</td>
<td>-0.80</td>
</tr>
<tr>
<td></td>
<td>DY_t</td>
<td>-3.34</td>
<td>0.42</td>
<td>-4.53</td>
<td>-2.56</td>
<td>-0.87</td>
</tr>
</tbody>
</table>
CHAPTER 7. FORECASTING

Figure 7.5: Plots of the time series of the logarithms of the equity premium, the dividend-price ratio and dividend yield.

Bill rate as the risk-free rate $R_{f_t}$ and the annual market return on the S&P 500 index $r_{mt}$. Figure 7.5 provides plots of the time series of the logarithms of the equity premium, dividend yield, and dividend-price ratio.

Two predictive regressions for $EQP_t$ are considered, namely,

$EQP_t = \alpha_p + \beta_p DP_{t-1} + \nu_{pt}$  \hspace{1cm} (7.20)

$EQP_t = \alpha_y + \beta_y DY_{t-1} + \nu_{yt}$. \hspace{1cm} (7.21)

The parameter estimates obtained from estimating these equations for two different sample periods (1926 to 1990 and 1926 to 2002) are reported in Table 7.3.

These results suggest that the lagged dividend price ratio, $DP_{t-1}$, and the lagged dividend yield, $DY_{t-1}$, have some forecasting power for the equity premium over the period 1926 - 1990, at least when $EQP_t$ is defined using the S&P 500 index. It is notable, however, that the size of both slope coefficients are substantially reduced when the sample size is increased to 2002.
7.8. PREDICTING THE EQUITY PREMIUM

Predictive regressions for the equity premium using the dividend price ratio, $d_t - p_t$, and the dividend yield, $d_t - p_{t-1}$, as explanatory variable predictors. Figures in parentheses are $t$ statistics.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$R^2$</th>
<th>$R^2$ Std. error</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{Pt-1}$</td>
<td>0.612</td>
<td>0.176</td>
<td>0.058</td>
<td>0.043</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>(2.146)</td>
<td>(1.959)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DY_{t-1}$</td>
<td>0.898</td>
<td>0.270</td>
<td>0.104</td>
<td>0.090</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>(2.851)</td>
<td>(2.683)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample 1926 - 1990

<table>
<thead>
<tr>
<th>Predictor</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$R^2$</th>
<th>$R^2$ Std. error</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{Pt-1}$</td>
<td>0.390</td>
<td>0.102</td>
<td>0.044</td>
<td>0.031</td>
<td>0.198</td>
</tr>
<tr>
<td></td>
<td>(2.128)</td>
<td>(1.839)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$DY_{t-1}$</td>
<td>0.504</td>
<td>0.138</td>
<td>0.070</td>
<td>0.058</td>
<td>0.195</td>
</tr>
<tr>
<td></td>
<td>(2.639)</td>
<td>(2.364)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample 1926 - 2002

The sub-sample instability of the estimated regression coefficients in Table 7.3 is further illustrated by considering the recursive plots of the slope coefficients from equations (7.21) and (7.20). Figure 7.6 reveals that the coefficient on $D_{Pt-1}$ increases over time while the coefficient on $DY_{t-1}$ steadily decreases. In other words, as time progresses forecasts appear to rely less on $DY_{t-1}$ and more on $D_{Pt-1}$ despite the fact that the coefficient on $DY_{t-1}$ appears more reliable in terms of statistical significance measured by the size of its $t$ ratio.

The main tool for interpreting the performance of the predictive regressions used by Goyal and Welch is a plot of the cumulative sum of squared one-step-ahead forecast errors of the predictive regressions expressed relative to the forecast error of the best current estimate of the mean of the equity premium. Let the one-step-ahead forecast errors of the dividend yield and dividend price ratio models be $\tilde{v}_{yt}$ and $\tilde{v}_{pt}$, respectively, and let the forecast errors for the unconditional mean estimate be $\tilde{u}_t$ which is simply the centered value of $EQP_t$. Figure 7.7 plots the two series

$$SSE(p) = \sum_{t=1946}^{2002} (\tilde{u}_t^2 - \tilde{v}_{pt}^2)$$ [Dividend-Price Ratio Model]

$$SSE(y) = \sum_{t=1946}^{2002} (\tilde{u}_t^2 - \tilde{v}_{yt}^2).$$ [Dividend Yield Model]

A positive value for $SSE$ means that the model forecasts are superior to the forecasts based solely on the mean thus far. A positive slope implies that over the recent year the forecasting model performs better than the mean.
Figure 7.6: Recursive estimates of the slope coefficients on the dividend-price ratio and the dividend yield from (7.20) and (7.21), respectively, computed for the period 1940 to 2002. The grey area denotes a one standard deviation band around the estimate.

Figure 7.7 indicates that the forecasting ability of a predictive regression using the dividend yield is poor because $SSE(y)$ is almost uniformly less than zero. There are two main years in the mid-1970s and again around 2000 when $SSE(y)$ has a positive slope (yet still negative) but these episodes are exceptions. The forecasting performance of the predictive regression using the dividend price ratio is slightly better than the forecasts generated by the mean as the summary measure $SSE(p) > 0$. This conclusion is supported by Figure 7.6 which shows the slope coefficient and its one standard deviation bands to be above the origin, albeit close to the origin over much of the period.

Table 7.4 provides forecast evaluation statistics for forecasts of the equity premium that are based on an estimation sample over 1926 - 1992 and a forecast
7.8. PREDICTING THE EQUITY PREMIUM

Figure 7.7: Plots of the cumulative sum of squared relative one-step-ahead forecast errors obtained from the equity premium predictive regressions for the period 1946 to 2002. The squared one-step-ahead forecast errors obtained from the models are subtracted from the squared one-step-ahead forecast errors based solely on the best current estimate of the unconditional mean of the equity premium.

The results in Table 7.4 indicate that the dividend price ratio predictor of the equity premium unambiguously provides the better forecast. The two combination forecasts provide only slender support for using forecast combinations in this illustration, these forecasts being dominated by the dividend price predictor. The simplicity of this forecasting problem, the small number of competing models, the fact that the dividend price ratio predictor dominates the dividend yield predictor, and the relative short horizon all make conditions less favourable for a successful combination forecast. In situations
Table 7.4

Performance of forecasts of the equity premium obtained from the dividend-price ratio model, the dividend yield model, a simple combination forecast, and a combination forecast that uses optimal weights. The estimation period is 1926-1992 and the forecast period is 1993-2002.

<table>
<thead>
<tr>
<th>Forecast Statistic</th>
<th>Dividend Price</th>
<th>Dividend Yield</th>
<th>Simple Combination</th>
<th>Optimal Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE</td>
<td>0.1614</td>
<td>0.1875</td>
<td>0.1701</td>
<td>0.1676</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.9531</td>
<td>1.1621</td>
<td>1.0067</td>
<td>0.9840</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0392</td>
<td>0.0525</td>
<td>0.0444</td>
<td>0.0434</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.1980</td>
<td>0.2292</td>
<td>0.2107</td>
<td>0.2085</td>
</tr>
</tbody>
</table>

that involve many competing forecast models, each with its own potential advantages, and in forecast exercises over longer horizons where successful forecasting is inevitably more challenging, there is much greater potential for successful implementation of forecast combination.

There are some important practical lessons to learn from predictive regressions. The first is that good in-sample performance does not necessarily imply that a fitted regression equation will provide good forecasts. Parameter instability and structural change are inevitable realities that challenge good predictive performance. Second, there are major conceptual and technical difficulties that arise from differences in the time series properties of explanatory predictors and dependent variables. As in the above illustrations, the dependent variable of interest in forecasting exercises is often a stationary variable such as the equity premium which is difficult to predict well, as witnessed by the low $R^2$ statistics and poor predictive capability of the regressors demonstrated by the findings in Table 7.4. Furthermore, the explanatory predictors that are chosen for such regressions as being the most relevant variables are often time series whose autoregressive roots are typically near unity, signifying near nonstationarity.

Regression equations that relate such variables potentially suffer from an imbalance in the time series properties of the explanatory predictors and the dependent variable. For instance, Stambaugh (1999) found that dividend ratios have time series behaviour similar to random walks whereas equity premia are typically stationary. Dividend ratios may then be good predictors of their own future behaviour but may be viewed as only marginally relevant predictors of the future path of the equity premium. Therefore the coefficients in such imbalanced regressions must naturally be expected to be small and the explanatory power, as well as predictive power, of the regressions is similarly expected to be slight. Since much of the data that are employed in financial predictive regression exercises have nonstationary characteristics, these complications present major challenges for forecasters and active research on this
7.9 Stochastic Simulation of Value at Risk

Forecasting need not necessarily be about point forecasts or best guesses. Sometimes important information is conveyed by the degree of uncertainty inherent in the best guess. One important application of this uncertainty is the concept of value at risk which was introduced in Chapter 2. Stated formally, Value-at-Risk represents the losses that are expected to occur with probability $\alpha$ on an asset or portfolio of assets, $P$, after $N$ days. The $N$-day $(1 - \alpha)\%$ value at risk is expressed as $\text{VaR} (P, N, 1 - \alpha)$.

Recall from Chapter 2 that value at risk may be computed by historical simulation, the variance-covariance method, or Monte Carlo simulation. Using a model to make forecasts of future values of the asset or portfolio and then assessing the uncertainty in the forecast is the method of Monte Carlo simulation. In general, simulation refers to any method that randomly generates repeated trials of a model and seeks to summarise uncertainty in the model forecast in terms of the distribution of these random trials. The steps to perform a simulation for a simple AR(1) model are as follows:

**Step 1: Estimate the model**

Given observations $y_t = \{y_1, \cdots, y_T\}$, estimate the proposed model

$$y_t = \phi_0 + \phi_1 y_{t-1} + \nu_t,$$

and compute and store the residuals, $\tilde{\nu}_t$.

**Step 2: Simulate the model**

Now use the model to make a forecast but instead of making a single forecast based on a best guess for the unknowns, make explicit allowance for uncertainty by including a disturbance term. This disturbance term may be obtained either by drawing from some assumed distribution (such as the normal distribution) or by taking a random draw from the residuals, $\tilde{\nu}_t$, computed in Step 1. The forecasts are constructed as follows:

$$\hat{y}_{T+1}^1 = \phi_0 + \phi_1 y_T + \tilde{\nu}_{T+1},$$
$$\hat{y}_{T+2}^1 = \phi_0 + \phi_1 \hat{y}_{T+1} + \tilde{\nu}_{T+2},$$
$$\vdots$$
$$\hat{y}_{T+H}^1 = \phi_0 + \phi_1 \hat{y}_{T+H-1} + \tilde{\nu}_{T+H},$$

where $\tilde{\nu}_{T+H}$ represents a random draw from $\tilde{\nu}_t$. 

Step 3: Repeat
Step 2 is now repeated $S$ times to obtain

$$
\hat{y}_{T+1}^1, \hat{y}_{T+1}^2, \hat{y}_{T+1}^3, \ldots, \hat{y}_{T+S-1}^1, \hat{y}_{T+S-1}^2, \hat{y}_{T+S}^2, \\
\hat{y}_{T+2}, \hat{y}_{T+2}^2, \hat{y}_{T+2}^3, \ldots, \hat{y}_{T+S+1}^1, \hat{y}_{T+S+1}^2, \hat{y}_{T+S+2}^2, \\
\vdots, \vdots, \hat{y}_{T+H}^1, \hat{y}_{T+H}^2, \hat{y}_{T+H}^3, \ldots, \hat{y}_{T+S+H}^1, \hat{y}_{T+S+H}^2, \hat{y}_{T+S+H}^2.
$$

Step 4: Summarise the uncertainty
Each column of this ensemble of forecasts is a possible outcome of the model and therefore collectively the forecasts capture the uncertainty of the future value of $y_t$. In particular, the percentiles of these simulated forecasts for each time period $T + i$ provide an estimate of the distribution of the forecast at that time. The disturbances used to generate the forecasts are drawn from the actual one-step-ahead prediction errors and not from a normal distribution and the forecast uncertainty will then reflect any non-symmetry or fat tails present in the estimated prediction errors.

Consider the case of United States monthly data on equity prices. Suppose that the asset in question is one which pays the value of the index. An investor who holds this asset in September 2016, the last date in the sample, would observe that the value of the portfolio is $2157.69. Suppose the investor wishes to know what the value of the asset will be in six months time in March 2017. It is not so much the best guess of the future value that is important, but rather it is the spread of the distribution of the forecast. The situation is illustrated in Figure 7.8 where the shaded region captures the 90% confidence interval of the forecast. Clearly, the investor needs to take this spread of likely outcomes into account when thinking of the future value of the investment.
7.9. STOCHASTIC SIMULATION OF VALUE AT RISK

Figure 7.8: Simulation of the equity price index over the period July 2004 to December 2004. The ex ante forecasts are shown by the solid line while the confidence interval encapsulates the uncertainty inherent in the forecast.

Figure 7.9: Simulated distribution of the equity index and the profit/loss on the equity index over a six month horizon from July 2004.
Consider now the problem of computing the 99% Value-at-Risk for the asset which pays the value of the United States equity index over a time horizon of six months. On the assumption that equity returns are generated by an AR(1) model, the estimated equation is

\[ r_t = 0.2527 + 0.2839 r_{t-1} + \hat{\nu}_t, \]

which may be used to forecast returns for period \( T + 1 \) but ensuring that uncertainty is explicitly introduced. The forecasting equation is therefore

\[ \hat{r}_{T+1} = 0.2527 + 0.2839 r_T + \tilde{\hat{\nu}}_{T+1}, \]

where \( \tilde{\hat{\nu}}_{T+1} \) is a random draw from the computed one-step-ahead forecast errors computed by means of an in-sample static forecast. The value of the asset at \( T + 1 \) in repetition \( s \) is computed as

\[ \hat{P}^s_{T+1} = P_T \exp [\hat{r}_{T+1}/100], \]

where the forecast returns are adjusted so that they no longer expressed as percentages. A recursive procedure is now used to forecast the value of the asset out to \( T + 6 \) and the whole process is repeated \( S \) times. The distribution of the value of the asset at \( T + 6 \) after \( S \) replications is shown in Figure 7.9 based on the initial value at time \( T \) of \( P_T = $2157.69 \). The distribution of simulated losses obtained by subtracting the initial value of the asset from the terminal value is shown in Figure 7.9. The first percentile value of this distribution is \(-$675.97\) so that the six month 99% value at risk is \$675.97, where by convention the minus sign is dropped when reporting value at risk. Of course this approach is equally applicable to simulating value at risk for more complex portfolios comprising more than one asset and portfolios that include derivatives.

### 7.10 Exercises

The data required for the exercises are available for download as EViews work-files (*.wfl), Stata datafiles (*.dta), comma delimited text files (*.csv) and as Excel spreadsheets (*.xlsx).

1. **Ex Ante Forecasts of Equity Returns**

The data are monthly observations on United States equity prices and dividends for the period January 1871 to September 2016. Compute equity returns and dividend returns as

\[ re_t = 100 \times (\log P_t - \log P_{t-1}), \]
7.10. EXERCISES

\[ rd_t = 100 \times (\log D_t - \log D_{t-1}), \]

where \( P_t \) is the equity price index and \( D_t \) are dividend payments.

(a) Estimate an AR(1) model of equity returns, \( re_t \), with the sample period ending in September 2016. Generate forecasts of \( re_t \) from October to December 2016.

(b) Estimate an AR(2) model of equity returns, \( re_t \), with the sample period ending in September 2016. Generate forecasts of \( re_t \) from October to December 2016.

(c) Repeat parts (a) and (b) for dividend returns, \( rd_t \).

(d) Estimate a VAR(1) containing for \( re_t \) and \( rd_t \) with the sample period ending in September 2016. Generate forecasts of equity returns from October to December 2016.

(e) Estimate a VAR(2) for \( re_t \) and \( rd_t \) with the sample period ending in September 2016. Generate forecasts of equity returns from October to December 2016.

(f) Estimate a VECM(1) for the equity price, \( p_t \), dividend payments, \( d_t \), (where the lower cases denote logarithms) with the sample period ending in September 2016 and where the VECM specification contains a constant in both the VAR and the cointegrating equation. Generate forecasts of equity returns from October to December 2016.

(g) Repeat part (f) with the lag length in the VECM increasing from 1 to 2.

(h) Repeat part (g) with the VECM specification containing a constant in the cointegrating equation but not the VAR.

(i) Now estimate a VECM(1) containing the equity price, \( p_t \), dividend payments, \( d_t \), and earnings, \( e_t \), with the sample period ending in September 2016 and the specification as in part (f). Assume a cointegrating rank of 1. Generate forecasts of equity returns from October to December 2016.

(j) Repeat part (i) with the lag length in the VECM increasing from 1 to 2.

2. Ex Post Forecasts of Equity Returns

The data are the same monthly observations on United States equity prices and dividends for the period January 1871 to September 2016 as used in Exercise 1. Compute equity returns and dividend returns as

\[ re_t = 100 \times (\log P_t - \log P_{t-1}), \]
rd_{t} = 100 \times (\log D_{t} - \log D_{t-1}),

where \( P_{t} \) is the equity price index and \( D_{t} \) are dividend payments.

(a) Estimate an AR(1) model of equity returns, \( r_{et} \), with the sample period ending December 2015, and generate ex post forecasts from January to September 2016.

(b) Estimate a VAR(1) model of equity returns, \( r_{et} \) and dividend returns \( rd_{t} \) with the sample period ending December 2015, and generate ex post forecasts from January to September 2016.

(c) Estimate a VECM(1) model of the equity price, \( p_{t} \) and dividend payments \( d_{t} \) using a constant in both the cointegrating equation and the VAR, with the sample period ending December 2015. Generate ex post forecasts for equity prices and returns from January to September 2016.

(d) For each set of forecasts generated in parts (a) to (c), compute the MSE and the RMSE. Which is the better forecasting model? Discuss.

3. Regression Based Forecasts of Equity Returns

The data are the same monthly observations on United States equity prices and dividends for the period January 1871 to September 2016 as used in Exercise 1. Compute equity returns and dividend returns as

\[
re_{t} = 100 \times (\log P_{t} - \log P_{t-1}),
\]

\[
rd_{t} = 100 \times (\log D_{t} - \log D_{t-1}),
\]

where \( P_{t} \) is the equity price index and \( D_{t} \) are dividend payments.

(a) Estimate the following regression for a sample period ending in September 2016

\[
re_{t} = \beta_{1} + \beta_{2}rd_{t} + u_{t},
\]

Estimate an AR(1) model of dividend returns

\[
rd_{t} = \rho_{0} + \rho_{1}rd_{t-1} + v_{t},
\]

and combine this model with the estimated model in the first equation to generate forecasts of equity returns from October to December 2016.
7.10. EXERCISES

(b) Estimate an AR(2) model of dividend returns

\[ rd_t = \rho_0 + \rho_1 rd_{t-1} + \rho_2 rd_{t-2} + \nu_t, \]

and combine this model in place of the AR(1) model in part (a) to generate forecasts of equity returns from October to December 2016.

(c) Use the estimated model in part (a) to generate forecasts of equity returns from October to December 2016 assuming that dividend returns are:

(i) 3% per annum;
(ii) 10% per annum;
(iii) 3% per annum in October and 10% in November and December.

4. Pooling Forecasts

The data are daily percentage returns to seven hedge fund indices, from the 1st of April 2003 to the 28th of May 2010, a sample size of \( T = 1869 \).

(a) Estimate an AR(2) model of the returns on the equity market neutral hedge fund with the sample period ending on the 21st of May 2010 (Friday). Generate forecasts of \( y_{1t} \) for the next working week, from the 24th to the 28th of May, 2010.

(b) Repeat part (a) for S&P 500 returns.

(c) Estimate a VAR(2) containing the returns on the neutral hedge fund \( (y_{1t}) \) and the returns on the S&P 500 \( (y_{2t}) \), with the sample period ending on the 21st of May 2010 (Friday)

\[
\begin{align*}
y_{1t} &= \alpha_0 + \alpha_1 y_{1t-1} + \alpha_2 y_{1t-2} + \alpha_3 y_{2t-1} + \alpha_4 y_{2t-2} + \nu_{1t} \\
y_{2t} &= \beta_0 + \beta_1 y_{1t-1} + \beta_2 y_{1t-2} + \beta_3 y_{2t-1} + \beta_4 y_{2t-2} + \nu_{2t}.
\end{align*}
\]

Generate forecasts of \( y_{1t} \) for the next working week, from the 24th to the 28th of May, 2010.

(d) For the AR(2) and VAR(2) forecasts obtained for the returns on the neutral hedge fund \( (y_{1t}) \) and the S&P 500 \( (y_{2t}) \), compute the RMSE (a total of four RMSEs) for each model. Discuss which model yields the superior forecasts.
(e) Let \( f_{1t}^{AR} \) be the forecasts from the AR(2) model of the returns on the neutral hedge fund and \( f_{1t}^{VAR} \) be the corresponding VAR(2) forecasts. Restricting the sample period just to the forecast period, 24th to the 28th of May, estimate the following unrestricted regression which pools the two sets of forecasts

\[
y_{1t} = \omega_0 + \omega_1 f_{1t}^{AR} + \omega_2 f_{1t}^{VAR} + \eta_t,
\]

where \( \eta_t \) is a disturbance term with zero mean and variance \( \sigma^2_{\eta} \).

Interpret the parameter estimates and discuss whether pooling the forecasts has improved the forecasts of the returns on the neutral hedge fund.

5. Evaluating Forecast Distributions

(a) (Correct Model Specification) Simulate \( y_1, y_2, \ldots, y_{1000} \) observations \((T = 1000)\) from the true model given by a \( N(0,1) \) distribution. Assuming that the specified model is also \( N(0,1) \), for each \( t \) compute the probability integral transform

\[
u_t = \Phi(y_t).
\]

Interpret the properties of the histogram of \( u_t \).

(b) (Mean Misspecification) Repeat part (a) except that the true model is \( N(0.5,1) \) and the misspecified model is \( N(0,1) \).

(c) (Variance Misspecification) Repeat part (a) except that the true model is \( N(0,2) \) and the misspecified model is \( N(0,1) \).

(d) (Skewness Misspecification) Repeat part (a) except that the true model is the standardised gamma distribution

\[
y_t = \frac{g_t - b a}{\sqrt{b^2 r}},
\]

where \( g_t \) is a gamma random variable with parameters \( \{b = 0.5, a = 2\} \) and the misspecified model is \( N(0,1) \).

(e) (Kurtosis Misspecification) Repeat part (a) except that the true model is the standardised Student t distribution

\[
y_t = \frac{s_t}{\sqrt{v - 2}},
\]

where \( s_t \) is a Student t random variable with degrees of freedom equal to \( v = 5 \), and the misspecified model is \( N(0,1) \).

6. Predicting the Equity Premium
7.10. EXERCISES

The data are annual observations on the return to the S&P 500 Index, \( r_{mt} \), equity premium, \( EQP_t \), the dividend price ratio, \( DP_t \), and the dividend yield \( DY_t \). The data are identical to those used by Goyal and Welch (2003) in their research on the determinants of the United States equity premium.

(a) Compute basic summary statistics for the market return, the equity premium, the dividend-price ratio and the dividend yield.
(b) Plot these variables and compare the results with Figure 7.5.
(c) Estimate the predictive regressions

\[
EQP_t = \alpha_p + \beta_p DP_{t-1} + \nu_p t
\]
\[
EQP_t = \alpha_y + \beta_y DY_{t-1} + \nu_y t,
\]

for two different sample periods, 1926 to 1990 and 1926 to 2002, and compare your results with Table 7.3.
(d) Estimate the regressions recursively using data up to 1940 as the starting sample in order to obtain recursive estimates of \( \beta_y \) and \( \beta_p \) together with 95% confidence intervals. Plot and interpret the results.
(e) Estimate the predictive regressions using the sample period 1926 to 1990. Use these models to provide ex post forecasts of \( EQP_t \) over the period 1993 to 2002. Evaluate these forecasts using commonly used forecast evaluation statistics and comment on your results.

7. Simulating VaR for a Single Asset

The data are monthly observations on United States equity prices from January 1871 to September 2016. Equity returns expressed as a percentage, are given by

\[
re_t = 100 \times (\log P_t - \log P_{t-1}).
\]

The aim is to compute the six-month 99% Value-at-Risk for an asset which pays the value of the index.

(a) Assume that the equity returns are generated by an AR(1) model

\[
re_t = \phi_0 + \phi_1 re_{t-1} + \nu_t.
\]

Estimate the model and compute the residuals, \( \tilde{\nu}_t \).
(b) Generate 1000 forecasts of the terminal equity price \( P_{T+6} \) using stochastic simulation by implementing the following steps.

i. Forecast \( \hat{r}e_{T+k}^g \) using the scheme

\[
\hat{r}e_{T+k}^g = \hat{\phi}_0 + \hat{\phi}_1 \hat{r}e_{T+k-1}^g + \tilde{\sigma}_{T+k},
\]

where \( \tilde{\sigma}_{T+k} \) is a random draw from the residuals, \( \tilde{\sigma}_t \).

ii. Compute the simulated equity price

\[
\hat{P}_{T+k}^g = \hat{P}_{T+k}^g \exp(\hat{r}e_{T+k}^g / 100).
\]

iii. Repeat (i) and (ii) for \( k = 1, 2, \cdots 6 \).

iv. Repeat (i), (ii) and (iii) for \( s = 1, 2, \cdots 1000 \).

(c) Compute the 99% Value-at-Risk based on the \( S \) simulated equity prices at \( T + 6 \) and the initial value of the index in September 2016.