Introductory Econometrics for Finance

Chris Brooks
The ISMA Centre, University of Reading
Contents

List of figures .......................... xii
List of tables .......................... xv
List of boxes .......................... xviii
List of screenshots ..................... xx
Preface .................................. xxi
Acknowledgements ..................... xxv

1 Introduction .......................... 1
1.1 What is econometrics? ............... 1
1.2 Is financial econometrics different from ‘economic econometrics’? Some stylised characteristics of financial data ...... 2
1.3 Types of data ........................ 4
1.4 Returns in financial modelling ...... 6
1.5 Steps involved in formulating an econometric model 8
1.6 Some points to consider when reading articles in the empirical financial literature 10
1.7 Outline of the remainder of this book 11

2 Econometric packages for modelling financial data 15
2.1 What packages are available? ...... 15
2.2 Choosing a package ................. 16
2.3 Accomplishing simple tasks using the two packages 17
2.4 WinRATS .......................... 18
2.5 EViews ............................ 31
2.6 Further reading ..................... 39
Appendix: economic software package suppliers 40

3 A brief overview of the classical linear regression model 42
3.1 What is a regression model? ...... 42
3.2 Regression versus correlation .... 43
3.3 Simple regression 43
3.4 Some further terminology 52
3.5 The assumptions underlying the classical linear regression model 55
3.6 Properties of the OLS estimator 56
3.7 Precision and standard errors 58
3.8 An introduction to statistical inference 64
3.9 Generalising the simple model to multiple linear regression 82
3.10 The constant term 83
3.11 How are the parameters (the elements of the $\beta$ vector) calculated in the generalised case? 85
3.12 A special type of hypothesis test: the $t$-ratio 88
3.13 Data mining and the true size of the test 89
3.14 An example of the use of a simple $t$-test to test a theory in finance: can US mutual funds beat the market? 90
3.15 Can UK unit trust managers beat the market? 93
3.16 The overreaction hypothesis and the UK stock market 95
3.17 Testing multiple hypotheses: the $F$-test 102
3.18 Sample EViews and RATS instructions and output for simple linear regression 108
Appendix: mathematical derivations of CLRM results 122
3A.1 Deriving the OLS coefficient estimator in the bivariate case 122
3A.2 Derivation of the OLS standard error estimators for the intercept and slope in the bivariate case 123
3A.3 Derivation of the OLS coefficient estimator in the multiple regression context 127
3A.4 Derivation of the OLS standard error estimator in the multiple regression context 128

4 Further issues with the classical linear regression model 133
4.1 Goodness of fit statistics 133
4.2 Hedonic pricing models 139
4.3 Tests of non-nested hypotheses 142
4.4 Violations of the assumptions of the classical linear regression model 144
4.5 Assumption 1: $E(u_i) = 0$ 146
4.6 Assumption 2: $\text{var}(u_i) = \sigma^2 < \infty$ 147
4.7 Assumption 3: $\text{cov}(u_i, u_j) = 0$ for $i \neq j$ 155
4.8 Assumption 4: the $x_i$ are non-stochastic 178
Contents

4.9 Assumption 5: the disturbances are normally distributed 178
4.10 Multicollinearity 190
4.11 Adopting the wrong functional form 194
4.12 Omission of an important variable 197
4.13 Inclusion of an irrelevant variable 198
4.14 Parameter stability tests 198
4.15 A strategy for constructing econometric models and a discussion of model-building philosophies 208
4.16 Determinants of sovereign credit ratings 211
  Appendix: a brief introduction to principal components analysis 220
4A.1 An application of principal components to interest rates 222
4A.2 Calculating principal components in practice 225

5 Univariate time series modelling and forecasting 229
  5.1 Introduction 229
  5.2 Some notation and concepts 230
  5.3 Moving average processes 235
  5.4 Autoregressive processes 239
  5.5 The partial autocorrelation function 247
  5.6 ARMA processes 249
  5.7 Building ARMA models: the Box-Jenkins approach 255
  5.8 Example: constructing ARMA models in EViews 258
  5.9 Estimating ARMA models with RATS 268
  5.10 Examples of time series modelling in finance 272
  5.11 Exponential smoothing 275
  5.12 Forecasting in econometrics 277
  5.13 Forecasting using ARMA models in EViews 291
  5.14 Forecasting using ARMA models in RATS 293
  5.15 Estimating exponential smoothing models using EViews and RATS 295

6 Multivariate models 302
  6.1 Motivations 302
  6.2 Simultaneous equations bias 304
  6.3 So how can simultaneous equations models be validly estimated? 306
  6.4 Can the original coefficients be retrieved from the $\pi$s? 306
  6.5 Simultaneous equations in finance 309
  6.6 A definition of exogeneity 310
  6.7 A special case: a set of equations that looks like a simultaneous equations system, but isn’t 313
Contents

6.8 Estimation procedures for simultaneous equations systems 313
6.9 An application of a simultaneous equations approach in finance: modelling bid–ask spreads and trading activity in the S&P 100 index options market 317
6.10 Simultaneous equations modelling using EViews and RATS 323
6.11 A Hausman test in RATS 328
6.12 Vector autoregressive models 330
6.13 Does the VAR include contemporaneous terms? 336
6.14 Block significance and causality tests 338
6.15 VARs with exogenous variables 340
6.16 Impulse responses and variance decompositions 340
6.17 An example of the use of VAR models: the interaction between property returns and the macroeconomy 343
6.18 VAR estimation in RATS and EViews 351

7 Modelling long-run relationships in finance 367
7.1 Stationarity and unit root testing 367
7.2 Testing for unit roots in EViews 383
7.3 Testing for unit roots in RATS 386
7.4 Cointegration 387
7.5 Equilibrium correction or error correction models 389
7.6 Testing for cointegration in regression: a residuals-based approach 391
7.7 Methods of parameter estimation in cointegrated systems 393
7.8 Lead-lag and long-term relationships between spot and futures markets 395
7.9 Testing for and estimating cointegrating systems using the Johansen technique based on VARs 403
7.10 Purchasing power parity 409
7.11 Cointegration between international bond markets 411
7.12 Testing the expectations hypothesis of the term structure of interest rates 418
7.13 Testing for cointegration and modelling cointegrated systems using EViews and RATS 420

8 Modelling volatility and correlation 437
8.1 Motivations: an excursion into non-linearity land 437
8.2 Models for volatility 441
8.3 Historical volatility 441
8.4 Implied volatility models 442
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.5 Exponentially weighted moving average models</td>
<td>442</td>
</tr>
<tr>
<td>8.6 Autoregressive volatility models</td>
<td>444</td>
</tr>
<tr>
<td>8.7 Autoregressive conditionally heteroscedastic (ARCH) models</td>
<td>445</td>
</tr>
<tr>
<td>8.8 Generalised ARCH (GARCH) models</td>
<td>452</td>
</tr>
<tr>
<td>8.9 Estimation of ARCH/GARCH models</td>
<td>455</td>
</tr>
<tr>
<td>8.10 Extensions to the basic GARCH model</td>
<td>468</td>
</tr>
<tr>
<td>8.11 Asymmetric GARCH models</td>
<td>469</td>
</tr>
<tr>
<td>8.12 The GJR model</td>
<td>469</td>
</tr>
<tr>
<td>8.13 The EGARCH model</td>
<td>470</td>
</tr>
<tr>
<td>8.14 GJR and EGARCH in EViews</td>
<td>471</td>
</tr>
<tr>
<td>8.15 Estimating GJR and EGARCH models using RATS</td>
<td>472</td>
</tr>
<tr>
<td>8.16 Tests for asymmetries in volatility</td>
<td>474</td>
</tr>
<tr>
<td>8.17 GARCH-in-mean</td>
<td>480</td>
</tr>
<tr>
<td>8.18 Uses of GARCH-type models including volatility forecasting</td>
<td>482</td>
</tr>
<tr>
<td>8.19 Testing non-linear restrictions or testing hypotheses about non-linear models</td>
<td>490</td>
</tr>
<tr>
<td>8.20 Volatility forecasting: some examples and results from the literature</td>
<td>493</td>
</tr>
<tr>
<td>8.21 Stochastic volatility models revisited</td>
<td>501</td>
</tr>
<tr>
<td>8.22 Forecasting covariances and correlations</td>
<td>502</td>
</tr>
<tr>
<td>8.23 Covariance modelling and forecasting in finance: examples of model uses</td>
<td>503</td>
</tr>
<tr>
<td>8.24 Historical covariance and correlation</td>
<td>505</td>
</tr>
<tr>
<td>8.25 Implied covariance models</td>
<td>505</td>
</tr>
<tr>
<td>8.26 Exponentially weighted moving average models for covariances</td>
<td>506</td>
</tr>
<tr>
<td>8.27 Multivariate GARCH models</td>
<td>506</td>
</tr>
<tr>
<td>8.28 A multivariate GARCH model for the CAPM with time-varying covariances</td>
<td>510</td>
</tr>
<tr>
<td>8.29 Estimating a time-varying hedge ratio for FTSE stock index returns</td>
<td>512</td>
</tr>
<tr>
<td>8.30 Estimating multivariate GARCH models using RATS and EViews</td>
<td>516</td>
</tr>
<tr>
<td>Appendix: parameter estimation using maximum likelihood</td>
<td>526</td>
</tr>
<tr>
<td>9 Switching models</td>
<td>533</td>
</tr>
<tr>
<td>9.1 Motivations</td>
<td>533</td>
</tr>
<tr>
<td>9.2 Seasonalities in financial markets: introduction and literature review</td>
<td>536</td>
</tr>
</tbody>
</table>
## Contents

9.3 Modelling seasonality in financial data  537
9.4 Estimating simple piecewise linear functions  545
9.5 Markov switching models  546
9.6 An application of Markov switching models to the gilt–equity yield ratio  549
9.7 Estimation of Markov switching models in RATS  558
9.8 Threshold autoregressive models  559
9.9 Estimation of threshold autoregressive models  561
9.10 Specification tests in the context of Markov switching and threshold autoregressive models: a cautionary note  563
9.11 An example of applying a SETAR model to the French franc–German mark exchange rate  564
9.12 Threshold models and the dynamics of the FTSE 100 stock index and stock index futures market  567
9.13 A note on regime switching models and forecasting accuracy  571
9.14 Estimating threshold autoregressive models in RATS  571

### 10 Simulation methods

10.1 Motivations  577
10.2 Monte Carlo simulations  578
10.3 Variance reduction techniques  580
10.4 Bootstrapping  585
10.5 Random number generation  589
10.6 Disadvantages of the simulation approach to econometric or financial problem solving  590
10.7 An example of the use of Monte Carlo simulation in econometrics: deriving a set of critical values for a Dickey–Fuller test.  592
10.8 An example of how to simulate the price of a financial option  601
10.9 An example of the use of bootstrapping to calculate capital risk requirements  612

### 11 Conducting empirical research or doing a project or dissertation in finance

11.1 What is an empirical research project, and what is it for?  632
11.2 Selecting the topic  633
11.3 Working papers and literature on the Internet  636
11.4 Getting the data  636
11.5 Choice of computer software  639
Figures

1.1 Steps involved in forming an econometric model  
2.1 Sample time series plot produced by RATS  
2.2 Sample scatter plot produced by RATS  
3.1 Scatter plot of two variables, y and x  
3.2 Scatter plot of two variables with a line of best fit chosen by eye  
3.3 Method of OLS fitting a line to the data by minimising the sum of squared residuals  
3.4 Plot of a single observation, together with the line of best fit, the residual and the fitted value  
3.5 Scatter plot of excess returns on fund XXX versus excess returns on the market portfolio  
3.6 No observations close to the y-axis  
3.7 Effect on the standard errors of the coefficient estimates when $(x_t - \bar{x})$ are narrowly dispersed  
3.8 Effect on the standard errors of the coefficient estimates when $(x_t - \bar{x})$ are widely dispersed  
3.9 Effect on the standard errors of $x_t^2$ large  
3.10 Effect on the standard errors of $x_t^2$ small  
3.11 The normal distribution  
3.12 The t-distribution versus the normal  
3.13 Rejection regions for a two-sided 5% hypothesis test  
3.14 Rejection regions for a one-sided hypothesis test of the form $H_0: \beta = \beta^*, H_1: \beta > \beta^*$  
3.15 Rejection regions for a one-sided hypothesis test of the form $H_0: \beta = \beta^*, H_1: \beta < \beta^*$  
3.16 Critical values and rejection regions for a $t_{20,5\%}$  
3.17 Frequency distribution of t-ratios of mutual fund alphas (gross of transactions costs)
List of figures

3.18 Frequency distribution of t-ratios of mutual fund alphas (net of transactions costs) .................................................. 92
4.1 \( R^2 = 0 \) demonstrated by a flat estimated line, i.e. a zero slope coefficient ................................................................. 136
4.2 \( R^2 = 1 \) when all data points lie exactly on the estimated line ................................................................. 136
4.3 Effect of no intercept on a regression line ................................................................. 147
4.4 Graphical illustration of heteroscedasticity ................................................................. 148
4.5 Plot of \( \hat{u}_t \) against \( \hat{u}_{t-1} \), showing positive autocorrelation ................................................................. 157
4.6 Plot of \( \hat{u}_t \) over time, showing positive autocorrelation ................................................................. 157
4.7 Plot of \( \hat{u}_t \) against \( \hat{u}_{t-1} \), showing negative autocorrelation ................................................................. 158
4.8 Plot of \( \hat{u}_t \) over time, showing negative autocorrelation ................................................................. 159
4.9 Plot of \( \hat{u}_t \) against \( \hat{u}_{t-1} \), showing no autocorrelation ................................................................. 159
4.10 Plot of \( \hat{u}_t \) over time, showing no autocorrelation ................................................................. 160
4.11 A normal versus a skewed distribution ................................................................. 179
4.12 A leptokurtic versus a normal distribution ................................................................. 180
4.13 Regression residuals from stock return data, showing large outlier for October 1987 ................................................................. 183
4.14 Possible effect of an outlier on OLS estimation ................................................................. 185
4.15 Plot of a variable showing suggestion for break date ................................................................. 205
5.1 Autocorrelation function for sample MA(2) process ................................................................. 239
5.2 Sample autocorrelation and partial autocorrelation functions for an MA(1) model ................................................................. 251
5.3 Sample autocorrelation and partial autocorrelation functions for an MA(2) model ................................................................. 251
5.4 Sample autocorrelation and partial autocorrelation functions for a slowly decaying AR(1) model ................................................................. 252
5.5 Sample autocorrelation and partial autocorrelation functions for a more rapidly decaying AR(1) model ................................................................. 252
5.6 Sample autocorrelation and partial autocorrelation functions for a more rapidly decaying AR(1) model with negative coefficient ................................................................. 253
5.7 Sample autocorrelation and partial autocorrelation functions for a non-stationary model (i.e. a unit coefficient) ................................................................. 253
5.8 Sample autocorrelation and partial autocorrelation functions for an ARMA(1, 1) model ................................................................. 254
5.9 Use of an in-sample and an out-of-sample period for analysis ................................................................. 279
6.1 Impulse responses and standard error bands for innovations in unexpected inflation equation errors ................................................................. 350
List of figures

6.2 Impulse responses and standard error bands for innovations in the dividend yields 350
7.1 Value of $R^2$ for 1,000 sets of regressions of a non-stationary variable on another independent non-stationary variable 368
7.2 Value of $t$-ratio of slope coefficient for 1,000 sets of regressions for a non-stationary variable on another independent non-stationary variable 369
7.3 Example of a white noise process 373
7.4 Time series plot of a random walk versus a random walk with drift 374
7.5 Time series plot of a deterministic trend process 374
7.6 Autoregressive processes with differing values of $\phi (0,0.8,1)$ 375
8.2 The problem of local optima in maximum likelihood estimation 459
8.3 News impact curves for S&P 500 returns using coefficients implied from GARCH and GJR model estimates 479
8.4 Three approaches to hypothesis testing under maximum likelihood 491
8.5 Time-varying hedge ratios derived from symmetric and asymmetric BEKK models for FTSE returns 515
9.1 Sample time series plot illustrating a regime shift 534
9.2 Use of intercept dummy variables for quarterly data 539
9.3 Use of slope dummy variables 541
9.4 Piecewise linear model with threshold $x^*$ 547
9.5 Unconditional distribution of US GEYR together with a normal distribution with the same mean and variance 551
9.6 Value of GEYR and probability that it is in the high GEYR regime for the UK 553
9.7 Mean–variance efficient frontier for the UK 555
Tables

2.1 Econometric software packages  
3.1 Critical values from the standard normal versus t-distribution  
3.2 Classifying hypothesis testing errors and correct conclusions  
3.3 Summary statistics for the estimated regression results for (3.52)  
3.4 Summary statistics for unit trust returns, January 1979–May 2000  
3.5 CAPM regression results for unit trust returns, January 1979–May 2000  
3.6 Is there an overreaction effect in the UK stock market?  
3.7 Overreaction versus small firm effects for a 1-year horizon  
3.8 Overreaction versus small firm effects for a 2-year horizon  
3.9 Overreaction versus small firm effects for a 3-year horizon  
4.1 Hedonic model of rental values in Quebec City, 1990  
4.2 Constructing a series of lagged values and first differences  
4.3 Determinants and impacts of sovereign credit ratings  
4.4 Do ratings add to public information?  
4.5 What determines reactions to ratings announcements?  
4A.1 Principal component ordered eigenvalues for Dutch interest rates, 1962–1970  
4A.2 Factor loadings of the first and second principal components for Dutch interest rates, 1962–1970  
5.1 Uncovered interest parity tests results  
5.2 Forecast error aggregation  
6.1 Call bid–ask spread and trading volume regression  
6.2 Put bid–ask spread and trading volume regression  
6.3 Granger causality tests and implied restrictions on VAR models
List of tables

6.4 Marginal significance levels associated with joint $F$-tests that all 14 lags have no explanatory power for that particular equation in the VAR 347
6.5 Variance decompositions for the property sector index residuals 348
7.1 Critical values for DF tests 379
7.2 DF tests on log-prices and returns for high frequency FTSE data 397
7.3 Estimated potentially cointegrating equation and test for cointegration for high frequency FTSE data 398
7.4 Estimated error correction model for high frequency FTSE data 398
7.5 Comparison of out-of-sample forecasting accuracy 399
7.6 Trading profitability of the error correction model with cost of carry 401
7.7 Cointegration tests of PPP with European data 410
7.8 DF tests for international bond indices 412
7.9 Cointegration tests for pairs of international bond indices 413
7.10 Johansen tests for cointegration between international bond yields 414
7.11 Variance decompositions for VAR of international bond yields 415
7.12 Impulse responses for VAR of international bond yields 416
7.13 Tests of the expectations hypothesis using the US zero coupon yield curve with monthly data 420
8.1 GARCH versus implied volatility 496
8.2 EGARCH versus implied volatility 497
8.3 Out-of-sample predictive power for weekly volatility forecasts 499
8.4 Comparisons of the relative information content of out-of-sample volatility forecasts 501
8.5 Hedging effectiveness: summary statistics for portfolio returns 514
9.1 Values and significances of days of the week coefficients 540
9.2 Day-of-the-week effects with the inclusion of interactive dummy variables with the risk proxy 543
9.3 Estimated parameters for the Markov switching models 552
9.4 Average returns and volatility of returns for a bond, an equity and the Markov switching portfolios 555
9.5 SETAR model for FRF-DEM 565
9.6 FRF-DEM forecast accuracies 566
List of tables

9.7 Linear AR(3) model for the basis 569
9.8 A two-threshold SETAR model for the basis 570
10.1 Simulated Asian option price values 609
10.2 Autocorrelations of returns and their absolute values for currency futures 615
10.3 EGARCH estimates for currency futures returns 616
10.4 Autoregressive volatility estimates for currency futures returns 618
10.5 Minimum capital risk requirements for currency futures as a percentage of the initial value of the position 620
11.1 Journals in finance and econometrics 635
11.2 Useful Internet sites for financial literature 637
11.3 Suggested structure for a typical dissertation or project 640
A2.1 Normal critical values for different values of $\alpha$ 668
A2.2 Critical values of Student’s $t$-distribution for different probability levels, $\alpha$ and degrees of freedom, $v$ 669
A2.3 Upper 5% critical values for $F$-distribution 670
A2.4 Upper 1% critical values for $F$-distribution 671
A2.5 Chi-squared critical values for different values of $\alpha$ and degrees of freedom, $v$ 672
A2.6 Lower and upper 5% critical values for Durbin–Watson statistic 674
A2.7 Dickey–Fuller critical values for different significance levels, $\alpha$ 675
A2.8 Critical values for the Engle–Granger cointegration test on regression residuals with no constant in test regression 676
A2.9 Quantiles of the asymptotic distribution of the Johansen cointegration rank test statistics (constant in cointegrating vectors only) 677
A2.10 Quantiles of the asymptotic distribution of the Johansen cointegration rank test statistics (constant, i.e. a drift only in VAR and in cointegrating vector) 678
A2.11 Quantiles of the asymptotic distribution of the Johansen cointegration rank test statistics (constant in cointegrating vector and VAR, trend in cointegrating vector) 679
Boxes

1.1 The value of econometrics page 2
1.2 Time series data 4
1.3 Log returns 8
1.4 Points to consider when reading a published paper 11
2.1 WinRATS icon functions 20
2.2 Series organised by Observation 23
2.3 Series organised by Variables 23
2.4 Features of RATS 29
2.5 Features of EViews 38
3.1 Names of ys and xs in regression models 42
3.2 Reasons for the inclusion of the disturbance term 45
3.3 Assumptions concerning disturbance terms and their interpretation 56
3.4 Standard error estimators 60
3.5 Conducting a test of significance 70
3.6 Carrying out a hypothesis test using confidence intervals 74
3.7 The test of significance and confidence interval approaches compared 75
3.8 Type I and type II errors 79
3.9 Reasons for stock market overreactions 95
3.10 Ranking stocks and forming portfolios 96
3.11 Portfolio monitoring 97
3.12 Modelling the return on shares 118
4.1 Selecting between models 144
4.2 Conducting White’s test 148
4.3 ’Solutions’ for heteroscedasticity 152
4.4 Conditions for DW to be a valid test 164
4.5 Conducting a Breusch–Godfrey test 165
4.6 The Cochrane–Orcutt procedure 168
4.7 Observations for the dummy variable 183
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.8 Conducting a Chow test</td>
<td>199</td>
</tr>
<tr>
<td>5.1 Naive forecasting methods</td>
<td>282</td>
</tr>
<tr>
<td>6.1 Determining whether an equation is identified</td>
<td>307</td>
</tr>
<tr>
<td>6.2 Using the Hausman test</td>
<td>311</td>
</tr>
<tr>
<td>8.1 Testing for ‘ARCH effects’</td>
<td>448</td>
</tr>
<tr>
<td>8.2 Estimating an ARCH or GARCH model</td>
<td>456</td>
</tr>
<tr>
<td>8.3 Using maximum likelihood estimation in practice</td>
<td>459</td>
</tr>
<tr>
<td>10.1 Conducting a Monte Carlo simulation</td>
<td>579</td>
</tr>
<tr>
<td>10.2 Re-sampling the data</td>
<td>587</td>
</tr>
<tr>
<td>10.3 Re-sampling from the residuals</td>
<td>588</td>
</tr>
<tr>
<td>10.4 Using the Monte Carlo simulation: 1</td>
<td>593</td>
</tr>
<tr>
<td>10.5 Using the Monte Carlo simulation: 2</td>
<td>602</td>
</tr>
<tr>
<td>10.6 Generating draws from a GARCH process</td>
<td>603</td>
</tr>
</tbody>
</table>
Screenshots

2.1 The WinRATS screen display  page 19
2.2 Creating a workfile  32
2.3 An untitled workfile  32
2.4 Importing data into the workfile  33
2.5 The workfile containing loaded data  34
2.6 Verifying the data  34
3.1 Importing data from a CSV file  109
3.2 Scatterplot of two series  111
3.3 Scatterplot for spot and future series  116
4.1 Regression estimation window  154
4.2 Estimation options  155
4.3 Non-normality test results  181
4.4 Regression residuals, actual values and fitted series  186
4.5 Use of shading to identify an outlier  187
4.6 Shading used to find the exact date of the series break  187
4.7 Chow test for parameter stability  206
5.1 Estimating exponential smoothing models  295
6.1 Estimating the TOUCH equation  325
6.2 Estimating the TRADES equation  326
6.3 VAR inputs screen  358
6.4 VAR impulse responses  362
6.5 Combined response graphs  362
6.6 Variance decomposition graphs  363
7.1 Options menu for unit root tests  383
7.2 Residual plot to check for stationarity  422
7.3 Johansen cointegration test  426
7.4 VAR specification for Johansen tests  429
8.1 Estimating a GARCH-type model  462
8.2 GARCH model estimation options  463
8.3 Forecasting from GARCH models  486
8.4 Running an EViews program  526
This chapter sets the scene for the book by discussing in broad terms the questions of what is econometrics, and what are the 'stylised facts' describing financial data that researchers in this area typically try to capture in their models. It also collects together a number of preliminary issues relating to the construction of econometric models in finance.

1.1 What is econometrics?

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

The list in box 1.1 is of course by no means exhaustive, but it hopefully gives some flavour of the usefulness of econometric tools in terms of their financial applicability.
1.2 Is financial econometrics different from ‘economic econometrics’? Some stylised characteristics of financial data

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

In economics, a serious problem is often a lack of data at hand for testing the theory or hypothesis of interest – this is often called a ‘small samples problem’. It might be, for example, that data are required on government budget deficits, or population figures, which are measured only on an annual basis. If the methods used to measure these quantities changed a quarter of a century ago, then only at most 25 of these annual observations are usefully available.
Two other problems that are often encountered in conducting applied econometric work in the arena of economics are those of measurement error and data revisions. These difficulties are simply that the data may be estimated, or measured with error, and will often be subject to several vintages of subsequent revisions. For example, a researcher may estimate an economic model of the effect on national output of investment in computer technology using a set of published data, only to find that the data for the last two years have been revised substantially in the next, updated publication.

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

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

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

1.3 Types of data

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

1.3.1 Time series data

Time series data, as the name suggests, are data that have been collected over a period of time on one or more variables. Time series data have associated with them a particular frequency of observation or collection of data points. The frequency is simply a measure of the interval over, or the regularity with which, the data is collected or recorded. Box 1.2 shows some examples of time series data.

<table>
<thead>
<tr>
<th>Series</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industrial production</td>
<td>Monthly, or quarterly</td>
</tr>
<tr>
<td>Government budget deficit</td>
<td>Annually</td>
</tr>
<tr>
<td>Money supply</td>
<td>Weekly</td>
</tr>
<tr>
<td>The value of a stock</td>
<td>As transactions occur</td>
</tr>
</tbody>
</table>

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

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

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

Problems that could be tackled using time series data

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

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

1.3.2 Cross-sectional data

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

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

Problems that could be tackled using cross-sectional data

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

1.3.3 Panel data

Panel data have the dimensions of both time series and cross-sections, e.g. the daily prices of a number of blue chip stocks over two years. The estimation of panel regressions is an interesting and developing area, but will not be considered further. Interested readers are referred to the excellent text by Baltagi (1995).
Fortunately, virtually all of the standard techniques and analysis in econometrics are equally valid for time series and cross-sectional data. This book will, however, concentrate mainly on time series data and applications since these are more prevalent in finance. For time series data, it is usual to denote the individual observation numbers using the index \( t \), and the total number of observations available for analysis by \( T \). For cross-sectional data, the individual observation numbers are indicated using the index \( i \), and the total number of observations available for analysis by \( N \). Note that there is, in contrast to the time series case, no natural ordering of the observations in a cross-sectional sample. For example, the observations \( i \) might be on the price of bonds of different firms at a particular point in time, ordered alphabetically by company name. So, in the case of cross-sectional data, there is unlikely to be any useful information contained in the fact that Northern Rock follows National Westminster in a sample of UK bank credit ratings, since it is purely by chance that their names both begin with the letter ‘N’. On the other hand, in a time series context, the ordering of the data is relevant since the data are usually ordered chronologically.

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

### 1.4 Returns in financial modelling

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

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

**Simple returns**  
\[ R_t = \frac{p_t - p_{t-1}}{p_{t-1}} \times 100\% \quad (1.1) \]

**Continuously compounded returns**  
\[ r_t = 100\% \times \ln \left( \frac{p_t}{p_{t-1}} \right) \quad (1.2) \]

where:  
- \( R_t \) denotes the simple return at time \( t \)  
- \( r_t \) denotes the continuously compounded return at time \( t \)  
- \( p_t \) denotes the asset price at time \( t \)  
- \( \ln \) denotes the natural logarithm

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

Alternatively, it is often assumed that stock price time series have been adjusted and the dividends added back in to generate a total return index. Returns generated using either of the two formulae presented above thus prove a measure of the total return that would accrue to a holder of the asset during time \( t \).

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

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

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

But this does not work for the continuously compounded returns, so that they are not additive across a portfolio. The
Box 1.3 Log returns

1 Log-returns have the nice property that they can be interpreted as continuously compounded returns – so that the frequency of compounding of the return does not matter and thus returns across assets can more easily be compared.

2 Continuously compounded returns are time-additive. For example, suppose that a weekly returns series is required and daily log returns have been calculated for 5 days, numbered 1 to 5, representing the returns on Monday through Friday. It is valid to simply add up the 5 daily returns to obtain the return for the whole week:

\[
\begin{align*}
\text{Monday return} & \quad r_1 = \ln \left( \frac{p_1}{p_0} \right) = \ln p_1 - \ln p_0 \quad (1.4) \\
\text{Tuesday return} & \quad r_2 = \ln \left( \frac{p_2}{p_1} \right) = \ln p_2 - \ln p_1 \quad (1.5) \\
\text{Wednesday return} & \quad r_3 = \ln \left( \frac{p_3}{p_2} \right) = \ln p_3 - \ln p_2 \quad (1.6) \\
\text{Thursday return} & \quad r_4 = \ln \left( \frac{p_4}{p_3} \right) = \ln p_4 - \ln p_3 \quad (1.7) \\
\text{Friday return} & \quad r_5 = \ln \left( \frac{p_5}{p_4} \right) = \ln p_5 - \ln p_4 \quad (1.8) \\
\text{Return over the week} & \quad \ln p_5 - \ln p_0 = \ln \left( \frac{p_5}{p_0} \right) \quad (1.9)
\end{align*}
\]

The fundamental reason why this is the case is that the log of a sum is not the same as the sum of a log, since the operation of taking a log constitutes a non-linear transformation. Calculating portfolio returns in this context must be conducted by first estimating the value of the portfolio at each time period and then determining the returns.

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

1.5 Steps involved in formulating an econometric model

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

The steps involved in the model construction process are now listed and described. Further details on each stage are given in subsequent chapters of this book.
Figure 1.1
Steps involved in forming an econometric model

1a. Economic or financial theory (previous studies) → 1b. Formulation of an estimable theoretical model → 2. Collection of data → 3. Model estimation → 4. Is the model statistically adequate? 
   - No → Reformulate model 
   - Yes → 5. Interpret model → 6. Use for analysis

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

- **Step 2: collection of data relevant to the model** The data required may be available electronically through a financial information provider, such as Reuters, Bridge Telerate, or Primark Datastream, or from published government figures. Alternatively, the required data may be available only via a survey after distributing a set of questionnaires.

- **Step 3: choice of estimation method relevant to the model proposed in step 1** For example, is a single equation or multiple equation technique to be used?

- **Step 4: statistical evaluation of the model** What assumptions were required to estimate the parameters of the model optimally? Were these assumptions satisfied by the data or the model? Also, does the model adequately describe the data? If the answer is ‘yes’, proceed to step 5; if not, go back to steps 1–3 and either reformulate the model, collect more data, or select a different estimation technique that has less stringent requirements.

- **Step 5: evaluation of the model from a theoretical perspective** Are the parameter estimates of the sizes and signs that the theory or
intuition from step 1 suggested? If the answer is ‘yes’, proceed to step 6; if not, again return to stages 1–3.

- **Step 6: use of model**  
  When a researcher is finally satisfied with the model, it can then be used for testing the theory specified in step 1, or for formulating forecasts or suggested courses of action. This suggested course of action might be for an individual (e.g. ‘if inflation and GDP rise, buy stocks in sector X’), or as an input to government policy (e.g. ‘when equity markets fall, program trading causes excessive volatility and so should be banned’).

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

### 1.6 Some points to consider when reading articles in the empirical finance literature

As stated above, one of the defining features of this book relative to others in the area is in its use of published academic research as examples of the use of the various techniques. The papers examined in this book have been chosen for a number of reasons. Above all, they represent in this author’s opinion a clear and specific application in finance of the techniques covered in this book. They were also required to be published in a peer-reviewed journal, and hence to be widely available.

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

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

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

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

### 1.7 Outline of the remainder of this book

**Chapter 2**

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

**Chapter 3**

This introduces the classical linear regression model (CLRM). The ordinary least squares (OLS) estimator is derived and its interpretation discussed. The conditions for OLS optimality are stated and explained. Single and multiple hypothesis testing
frameworks are developed and examined in the context of the linear model. Examples employed include tests of the ‘overreaction hypothesis’ in the context of the UK stock market.

**Chapter 4**

This continues and develops the material of chapter 3 to consider goodness of fit statistics, diagnostic testing and the consequences of violations of the CLRM assumptions, along with plausible remedial steps. Model-building philosophies are discussed with particular reference to the general-to-specific approach. Finally, the main principles of principal components analysis are briefly discussed in an appendix. Applications covered in this chapter include hedonic models of rental values and the determination of sovereign credit ratings.

**Chapter 5**

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

**Chapter 6**

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

**Chapter 7**
The first section of the chapter discusses unit root processes and presents tests for non-stationarity in time series. The concept of and tests for cointegration, and the formulation of error correction models, are then discussed in the context of both the univariate framework of Engle–Granger, and the multivariate framework of Johansen. Applications studied in chapter 7 include spot and futures markets, tests for cointegration between international bond markets and tests of the purchasing power parity (ppp) exchange rates hypothesis.

**Chapter 8**
This covers the highly popular topic of volatility and correlation modelling and forecasting. This chapter starts by discussing in general terms the issue of non-linearity in financial time series. The class of ARCH (AutoRegressive Conditionally Heteroscedastic) models and the motivation for this formulation are then discussed. Other models are also presented, including extensions of the basic model such as GARCH, GARCH-M, EGARCH and GJR formulations. Examples of the huge number of applications are discussed, with particular reference to stock returns. Multivariate GARCH models are described, and applications to the estimation of conditional betas and time-varying hedge ratios, and to financial risk measurement, are given.

**Chapter 9**
This discusses testing for and modelling regime shifts or switches of behaviour in financial series that can arise from changes in government policy, market trading conditions or microstructure changes, among other causes. This chapter introduces the Markov switching approach to dealing with regime shifts. Threshold autoregression is also discussed, along with issues relating to the estimation of such models. Examples include the modelling of
exchange rates within a managed floating environment, modelling and forecasting the gilt–equity yield ratio, and models of movements of the difference between spot and futures prices.

Chapter 10
This presents an introduction to what is arguably one of the most rapidly developing areas in financial modelling: that of simulations. Motivations are given for the use of repeated sampling, and a distinction is drawn between Monte Carlo simulation and bootstrapping. The reader is shown how to set up a simulation, and examples are given in options pricing and financial risk management to demonstrate the usefulness of these techniques.

Chapter 11
This offers suggestions related to conducting a project or dissertation in empirical finance. It introduces the sources of financial and economic data available on the Internet and elsewhere, and recommends relevant online information and literature on research in financial markets and financial time series. The chapter also suggests ideas for what might constitute a good structure for a dissertation on this subject, how to generate ideas for a suitable topic, what format the report could take, and some common pitfalls.

Chapter 12
This summarises the book and concludes. Several recent developments in the field, which are not covered elsewhere in the book, are also mentioned. Some tentative suggestions for possible growth areas in the modelling of financial time series are also given.