

ANALYSIS OF FINANCIAL DATA

by

Gary Koop

University of Strathclyde



John Wiley & Sons, Ltd

Contents

Preface		ix
Chapter 1	Introduction	1
	Organization of the book	3
	Useful background	4
	Appendix 1.1: Concepts in mathematics used in this book	4
Chapter 2	Basic data handling	9
	Types of financial data	9
	Obtaining data	15
	Working with data: graphical methods	16
	Working with data: descriptive statistics	21
	Expected values and variances	24
	Chapter summary	26
	Appendix 2.1: Index numbers	27
	Appendix 2.2: Advanced descriptive statistics	30
Chapter 3	Correlation	33
	Understanding correlation	33
	Understanding why variables are correlated	39
	Understanding correlation through XY -plots	40
	Correlation between several variables	44
	Covariances and population correlations	45
	Chapter summary	47
	Appendix 3.1: Mathematical details	47

Chapter 4	An introduction to simple regression	49
	Regression as a best fitting line	50
	Interpreting OLS estimates	53
	Fitted values and R^2 : measuring the fit of a regression model	55
	Nonlinearity in regression	61
	Chapter summary	64
	Appendix 4.1: Mathematical details	65
Chapter 5	Statistical aspects of regression	69
	Which factors affect the accuracy of the estimate $\hat{\beta}$?	70
	Calculating a confidence interval for β	73
	Testing whether $\beta = 0$	79
	Hypothesis testing involving R^2 : the F -statistic	84
	Chapter summary	86
	Appendix 5.1: Using statistical tables for testing whether $\beta = 0$	87
Chapter 6	Multiple regression	91
	Regression as a best fitting line	93
	Ordinary least squares estimation of the multiple regression model	93
	Statistical aspects of multiple regression	94
	Interpreting OLS estimates	95
	Pitfalls of using simple regression in a multiple regression context	98
	Omitted variables bias	100
	Multicollinearity	102
	Chapter summary	105
	Appendix 6.1: Mathematical interpretation of regression coefficients	105
Chapter 7	Regression with dummy variables	109
	Simple regression with a dummy variable	112
	Multiple regression with dummy variables	114
	Multiple regression with both dummy and non-dummy explanatory variables	116
	Interacting dummy and non-dummy variables	120
	What if the dependent variable is a dummy?	121
	Chapter summary	122
Chapter 8	Regression with lagged explanatory variables	123
	Aside on lagged variables	125
	Aside on notation	127

	Selection of lag order	132
	Chapter summary	135
Chapter 9	Univariate time series analysis	137
	The autocorrelation function	140
	The autoregressive model for univariate time series	144
	Nonstationary versus stationary time series	146
	Extensions of the AR(1) model	149
	Testing in the AR(p) with deterministic trend model	152
	Chapter summary	158
	Appendix 9.1: Mathematical intuition for the AR(1) model	159
Chapter 10	Regression with time series variables	161
	Time series regression when X and Y are stationary	162
	Time series regression when Y and X have unit roots: spurious regression	167
	Time series regression when Y and X have unit roots: cointegration	167
	Time series regression when Y and X are cointegrated: the error correction model	174
	Time series regression when Y and X have unit roots but are not cointegrated	177
	Chapter summary	179
Chapter 11	Regression with time series variables with several equations	183
	Granger causality	184
	Vector autoregressions	190
	Chapter summary	203
	Appendix 11.1: Hypothesis tests involving more than one coefficient	204
	Appendix 11.2: Variance decompositions	207
Chapter 12	Financial volatility	211
	Volatility in asset prices: Introduction	212
	Autoregressive conditional heteroskedasticity (ARCH)	217
	Chapter summary	222
Appendix A	Writing an empirical project	223
	Description of a typical empirical project	223
	General considerations	225
Appendix B	Data directory	227
Index		231

Preface

This book aims to teach financial econometrics to students whose primary interest is not in econometrics. These are the students who simply want to apply financial econometric techniques sensibly in the context of real-world empirical problems. This book is aimed largely at undergraduates, for whom it can serve either as a stand-alone course in applied data analysis or as an accessible alternative to standard statistical or econometric textbooks. However, students in graduate economics and MBA programs requiring a crash-course in the basics of practical financial econometrics will also benefit from the simplicity of the book and its intuitive bent.

This book grew out of a previous book I wrote called *Analysis of Economic Data*. When writing my previous book I attempted to hold to the following principles:

1. It must cover most of the tools and models used in modern econometric research (e.g. correlation, regression and extensions for time series methods).
2. It must be largely non-mathematical, relying on verbal and graphical intuition.
3. It must contain extensive use of real data examples and involve students in hands-on computer work.
4. It must be short. After all, students in most degree programs must master a wide range of material. Students rarely have the time or the inclination to study statistics in depth.

In *Analysis of Financial Data* I have attempted to follow these principles as well but change the material so that it is of more interest for a financial audience. It aims to teach students reasonably sophisticated statistical tools, using simple non-mathematical intuition and practical examples. Its unifying themes are the related concepts of regression and correlation. These simple concepts are relatively easy to motivate using verbal and graphical intuition and underlie many of the sophisticated models (e.g. vector autoregressions and models of financial volatility such as ARCH

and GARCH) and techniques (e.g. cointegration and unit root tests) in financial research today. If a student understands the concepts of correlation and regression well, then she can understand and apply the techniques used in advanced financial econometrics and statistics.

This book has been designed for use in conjunction with a computer. I am convinced that practical hands-on computer experience, supplemented by formal lectures, is the best way for students to learn practical data analysis skills. Extensive problem sets are accompanied by different data sets in order to encourage students to work as much as possible with real-world data. Every theoretical point in the book is illustrated with practical financial examples that the student can replicate and extend using the computer. It is my strong belief that every hour a student spends in front of the computer is worth several hours spent in a lecture.

This book has been designed to be accessible to a variety of students, and thus, contains minimal mathematical content. Aside from some supplementary material in appendices, it assumes no mathematics beyond the pre-university level. For students unfamiliar with these basics (e.g. the equation of a straight line, the summation operator, logarithms), appendices at the end of chapters provide sufficient background.

I would like to thank my students and colleagues at the Universities of Edinburgh, Glasgow and Leicester for their comments and reactions to the lectures that formed the foundation of this book. Many reviewers also offered numerous helpful comments. Most of these were anonymous, but Ian Marsh, Denise Young, Craig Heinicke, Kai Li and Hiroyuki Kawakatsu offered numerous invaluable suggestions that were incorporated in the book. I am grateful, in particular, to Steve Hardman at John Wiley for the enthusiasm and expert editorial advice he gave throughout this project. I would also like to express my deepest gratitude to my wife, Lise, for the support and encouragement she provided while this book was in preparation.