# Preface

This book has its origin in a set of lecture notes prepared for a mixed group of graduates and advanced undergraduates at the London School of Economics. The graduate students come from a range of academic backgrounds, but most had taken mainstream econometrics courses using a text at the level of Johnston's (1972) *Econometric Methods* or Maddala's (1977) *Econometrics*. The undergraduates had also attended such a course, and most had taken a second-year course in probability and distribution theory.

The lectures evolved to try to meet the differing needs of these students. It was not possible to assume that all of them were rigorously drilled in the ramifications of the general linear model, though all should have been exposed to it, and some revision was deemed necessary. The degree of probabilistic intuition of the group also tended to vary widely. These topics needed to be reviewed with some care, but at the risk of boring the upper echelon of the class with repetition of familiar material. The compromise sought was to spend some time looking at old material in a new way. Rather than attempting to survey a wide range of advanced topics, the core of the course was an in-depth examination of the regression model. The written notes distributed to the class form the basis for several of the present chapters, although the opportunity has been taken to update the material, to treat some topics in more depth, and to add extra ones. Some of the additional material has been developed from lectures given for the M.Sc. Econometrics and Mathematical Economics at LSE, including the sections on nonlinear estimation, unit roots and cointegration. However, while the format of a book affords both author and reader more leisure to pursue the details than was possible in the lectures, the treatment remains self-contained and technically unified.

Something the lectures attempted, with what success the reader may judge, is to take the probabilistic foundations of the subject seriously. Given that economics is a non-experimental discipline, with nearly all observed variables being stochastically generated and not chosen by the investigator, the 'stochastic regressor' framework seems the appropriate one, and was adopted from the outset. Regression analysis is motivated by direct appeal to the notion of conditional expectation. One advantage of focusing on the conditioning approach to regression is that it neatly resolves an issue that often lands teachers and students in a tangle, that of correct and incorrect model specification. Students are traditionally told, for example, that regressors and disturbances should be uncorrelated for consistent estimation. This sounds sensible enough at first, but as one delves further into the analysis, puzzles creep in. The regressors are part of the data set, whereas the disturbances are not merely unobserved but are really fictional, having no existence outside the context of the model postulated. It is more satisfactory to focus on the properties of the conditional mean of a variable, deviations from which are orthogonal to the conditioning variables by construction. This leads naturally to the idea that every regression involves two sets of variables: the valid conditioning variables (those that may explain the regressand) and the explanatory variables (those that do explain it). The crucial step in model specification (defining the problem) is to specify the conditioning set. Model evaluation techniques can then be understood as guiding the choice of regressors from this set. Virtually the whole arsenal of regression diagnostics and modelling procedures can be accommodated in this basic framework. The disturbances would not even need to be mentioned to put the basic ideas across.

An attempt has also been made to unify the treatment of estimation theory through the notion of optimization estimators. A few simple results allow least squares, maximum likelihood, instrumental variables, GMM and other procedures to be viewed as different implementations of the same underlying principle. This approach saves a lot of duplication of effort in understanding the properties of the estimators, and yields results in many cases that are otherwise intractable. It also provides valuable insight, helping us to understand why, for example, an estimator described as 'maximum likelihood' can have good properties even when the data are not really distributed in a way that justifies the claim.

Thus, another aim has been to face up to the reality of partial specification and potential misspecification, which besets econometric practice. Students are often shown tightly specified models, involving features such as linearity, normality and independence of the disturbances, and stationarity of the time series composing the data set. While we all know that these are at best approximations to reality, there has been a pedagogic tendency to pile in all these conditions at the outset, and only subsequently (and probably sketchily) consider how to relax them. The inability of the classical statistical theory to handle data features such as dependence and heterogeneity is perhaps the chief reason why this approach is traditionally pursued. So far as is compatible with a reasonably straightforward treatment, the aim here has been to derive the theory of estimation and inference in the context of the minimum assumptions needed to justify it. For example, normality of the disturbances is not needed for most of the desirable asymptotic properties of the popular estimators (large sample efficiency being the chief exception). We also know that our models can sometimes throw up outliers in the residuals of a magnitude that sits unhappily with the normality assumption. Accordingly, this property is not emphasized: and the least squares estimator is not billed as maximum likelihood unless this assertion is specifically justified.

Likewise, the assumption of independence of disturbances really has no place in time series regression analysis. Almost all the economic time series we ever observe are serially dependent in some respect, and except in the context of joint (never mind conditional) normality, conditioning models do not induce independence of the disturbances. It must be imposed as an extra, arbitrary feature of the model, a *deus ex machina* that in any case has no essential role in the asymptotic analysis.

After the introductory chapters, independence is hardly used at all in the book except sometimes for illustration. Instead, appeal is made to the simple and very powerful concept of a martingale difference process, which has nearly all the desirable properties of an independent process, but a much better chance of being either generated in nature, or at worst, as the residuals from our modelling efforts.

This approach is made feasible by introducing some results from modern limit theory and, for this purpose, extensive reference is made at certain points to my book *Stochastic Limit Theory* (Davidson 1994a). There is a sense in which the two works can be viewed as companions, representing theory and application respectively. However, a knowledge of the mathematics used in *Stochastic Limit Theory* is not assumed here. Readers need to consult that book only to follow up the various proofs that are cited, and this is strictly optional.

There are always two sorts of question that an econometrics text must try to answer: the hows and the whys. The how questions (how to construct a model, and estimate it) are generally well covered by a range of existing texts. The why questions (why a particular technique is used in empirical work, or in other words, how its use can be justified) often receive less attention, because to answer them coherently requires the ground to be carefully prepared. This book does not try to answer all the how questions, and the coverage of models and methods is selective, although hopefully representative of current time series techniques. It does attempt to address some of the why questions seriously, which is the main excuse for the word 'theory' appearing in the title.

## An Overview of the Chapters

The book has been structured with a view to serving as the basis of a two-semester graduate course, and contains as much material as such a course could reasonably cover. The subject matter falls fairly naturally into four parts, and the chapters are grouped accordingly. Each part is reasonably self-contained, although making use of the preceding material. A one-semester course on regression analysis for time series might be based on Parts I and II (Chapters 1 to 8). These chapters deal with the classical linear regression model and its extension into time series analysis. No nonlinear analysis is attempted here, and the estimators are of the simplest kind, just ordinary least squares and two-stage least squares.

The first three chapters deal with the classical regression model specifically in the sense that the observations are treated as independently sampled. All the variables in the analysis are nonetheless treated as stochastic from the outset, so that notions of conditioning can be introduced. Chapter 1 introduces the main ideas and summarizes the matrix algebra, Chapter 2 explores the exact inference theory based on conditionally Gaussian disturbances (mostly identical to the fixed regressor theory) and Chapter 3 does the asymptotic version of the analysis, without the Gaussianity.

Part II (Chapters 4 to 8) is explicitly about time series analysis. The main idea developed here is that samples are not generally independent, but that sequential conditioning arguments yield a close analogue of the classical asymptotic analysis under independence. Chapters 4 and 5 survey the concepts of model building in time series, and the latter part of Chapter 5 is a digression into schemes for modelling phenomena such as lagged adjustment and rational expectations. Apart from their intrinsic interest, an important object of these sections is to motivate the later discussion of nonlinear modelling in time series. Chapter 6 develops the time series extensions of the asymptotic analysis, focusing for clarity on the estimation of the first-order autoregressive model. The key role of sequential conditioning is emphasized in the derivation of the properties of the estimators. Then, Chapter 7 looks at the asymptotic analysis of the multiple regression model, and goes on to deal with the issues of specification testing and model choice. Chapter 8 extends the analysis to simultaneous equations and thoroughly studies two-stage least squares. Although much of this material has a general application, the assumptions are once again relevant to the time series case.

Part III is an introduction to the general theory of optimization estimators (OEs, or M estimators) although, as in Part II, concepts of sequential conditioning are used throughout. Chapter 9 covers the main ideas, and Chapter 10 applies them to well-known problems, including nonlinear models, joint estimation of mean and variance, and misspecified models. Maximum likelihood estimation is deliberately covered, in Chapter 11, following the treatment of the general class of OEs in Chapters 9 and 10, so that readers can put this method into context and appreciate its special properties. This means that certain concepts are given a slightly unusual treatment. Something looking much like the Gaussian log-likelihood function for the linear model is analysed using OE theory, with no mention of Gaussian disturbances. The information matrix equality and likelihood ratio tests (which we generally refer to as 'analogue LR' tests) are among other topics treated in a novel manner, although hopefully one that clarifies the issues. Chapter 12 deals with the general theory of hypothesis testing, again in an OE framework. Emphasis is placed here on the role of the information matrix equality in simplifying test procedures, but also on the importance of robust procedures that do not depend on it. Chapter 13 covers classical system estimation, including the nonlinear case.

Finally, Part IV deals with unit roots and cointegration theory. This subject area has now become so large that a specialized text would be required to cover it comprehensively. A number of interesting topics have had to be given a cursory treatment. Rather, the aim has been to cover the basics thoroughly, and convey the flavour of the methods of analysis used in this field, especially the Brownian calculus. Chapter 14 discusses the probability essentials, and details the standard tests for unit roots. Chapter 15 looks at the estimation of cointegration models in a single-equation context. Chapter 16 deals with what is commonly known as Johansen's approach, and concludes the work with some considerations of structural cointegration modelling.

Although there is, unavoidably, a good deal of notation, the book aims as far as possible at a narrative and informal style. Proofs are set out formally in the text only if they are brief and essential to the development. Some longer and more technical proofs are segregated in chapter appendices, where they can be studied or ignored according to taste. The main appendices summarize the most important of the technical prerequisites, including matrix algebra and the elements of probability and distribution theory. Some readers may be surprised to find elementary material juxtaposed with some quite sophisticated arguments. This is deliberate. The book is intended to be accessible, at least in part, to students with no more mathematical training than accompanies the typical degree in economics, and things such students might not know are often pointed out. However, there is no attempt to censor difficult stuff where it is important to the story. In some books, technically advanced sections are 'starred' to warn off the timid reader, and this option was considered, but in the end it seemed better to let readers and teachers decide for themselves what was important and worth giving the effort to.

Literature references of two sorts are provided for the student who wishes to pursue the topics in greater depth. A list for 'Further Reading' is appended to each chapter, but with a few exceptions, these references are to textbooks, surveys and monographs, where the student might find more specialized coverage, or an alternative view. Most references to the journal literature, by contrast, are given in context.

## Acknowledgements

I must thank the people who, as they say, made this book possible. First and foremost, the remarkable people who taught me econometrics: Andy Chesher, Meghnad Desai, Jim Durbin, Terence Gorman, David Hendry, Grayham Mizon, the late Denis Sargan, and Ken Wallis. The econometrics group at LSE later provided a most congenial and stimulating working environment. Colleagues who at various times collaborated on the econometrics teaching, and participated in lively discussions at the Tuesday Econometrics workshops, have included Andrew Harvey, Javier Hidalgo, Jan Magnus, Stephen Pudney, Danny Quah, Peter Robinson, and Hugh Wills. Latterly my collaborator on various research projects, Robert de Jong, has been a source of indispensable insights on asymptotics. I am conscious of the debt I owe all the above-named and many others in developing the themes of this book, though needless to say, none are to blame for its undoubted shortcomings. The excellent students who attended the lectures contributed more than they could imagine. I don't just mean the ones who argued the case, pointed out errors and suggested improvements. Equally important were the majority whose enthusiasm and hard work made the enterprise worthwhile and enjoyable.

Blackwells commissioned the book longer ago than I now dare to admit, and the gestation process has proved tougher than I foresaw. Other projects have intervened, and for a while it seemed as though it might never be finished. The one consolation is that this one is a much better book than it would have been if delivered to the original schedule. Al Bruckner and Katie Byrne deserve credit for finally taking the thing by the scruff of the neck and driving it through to completion. It badly needed their encouragement, and I am grateful. Bruce Hansen, Len Gill, Brendan McCabe, and two anonymous referees have perused the manuscript at various stages, and I thank them for their helpful and encouraging comments.

For the typesetting design, created with the benefit of LaTeX, the author must

take any additional credit that may be going for himself. The original lecture notes and the first draft of the book were created in Chiwriter, but this word-processing package has since fallen by the wayside, and I must thank Jenny Firth for carefully retyping the manuscript in TEX using Scientific Word. The final version of the text has been developed from her files.

James Davidson

Cwmystwyth, Wales
September 1999