

## Preface

This book is designed to serve as the textbook for a first-year graduate course in econometrics. It has two distinguishing features. First, it covers a full range of techniques with the estimation method called the Generalized Method of Moments (GMM) as the organizing principle. I believe this unified approach is the most efficient way to cover the first-year materials in an accessible yet rigorous manner. Second, most chapters include a section examining in detail original applied articles from such diverse fields in economics as industrial organization, labor, finance, international, and macroeconomics. So the reader will know how to *use* the techniques covered in the chapter and under what conditions they are applicable.

Over the last several years, the lecture notes on which this book is based have been used at the University of Pennsylvania, Columbia University, Princeton University, the University of Tokyo, Boston College, Harvard University, and Ohio State University. Students seem to like the book a lot. My own experience from teaching out of the book is that students think the book is much better than the instructor.

### Prerequisites

The reader of this book is assumed to have a working knowledge of the basics of calculus, probability theory, and linear algebra. An understanding of the following concepts is taken for granted: functions of several variables, partial derivatives, integrals, random variables, joint distributions, independence, unconditional and conditional expectations, variances and covariances of vector random variables, normal distributions, chi-square distributions, matrix multiplication, inverses of matrices, the rank of a matrix, determinants, and positive definite matrices. Any relevant concepts above this level will be introduced as the discussion progresses. Results on partitioned matrices and Kronecker products are collected in the appendix. Prior exposure to undergraduate econometrics is not required.

### Organization of the Book

To understand how the book is organized, it is useful to distinguish between a *model* and an *estimation procedure*. The basic premise of econometrics is that economic data (such as postwar U.S. GDP) are random variables. A *model* is a

family of probability distributions that could possibly have generated the economic data. An *estimation procedure* is a data-based protocol for choosing from the model a particular distribution that is likely to have generated the data. Most estimation procedures in econometrics are a specialization of the GMM estimation principle. For example, when GMM is applied to a model called the classical linear regression model, the resulting estimation procedure is Ordinary Least Squares (OLS), the most basic estimation procedure in econometrics. This viewpoint is the organizing principle in the first six chapters of the book, where most of the standard estimation procedures are presented.

The book could have presented GMM in the first chapter, but that would deprive the reader of the option to follow a series of topics specific to OLS without getting distracted by GMM. For this reason I chose to use the first two chapters to present the finite-sample and large-sample theory of OLS. GMM is presented in Chapter 3 as a generalization of OLS.

A major expositional innovation of the book is to treat multiple-equation estimation procedures — such as Seemingly Unrelated Regressions (SUR), Three-Stage Least Squares (3SLS), the Random-Effects method, covered in Chapter 4, and the Fixed-Effects method covered in Chapter 5 — as special cases of the single-equation GMM of Chapter 3. This makes it possible to derive the statistical properties of those advanced techniques merely by suitably specializing the results about single-equation GMM developed in Chapter 3. Chapter 6 completes the book's discussion of GMM by indicating how serial dependence in the error term can be incorporated in GMM.

For some models in econometrics, Maximum Likelihood (ML) is the more natural estimation principle than GMM. ML is covered in Chapters 7 and 8. To make clear the relationship between GMM and ML, the book's discussion of ML starts out in Chapter 7 with an estimation principle called Extremum Estimators, which includes both ML and GMM as special cases. Applications of ML to various models are covered in Chapter 8.

The book also includes an extensive treatment of time-series analysis. Basic time-series topics are covered in Section 2.2 and in the first half of Chapter 6. That is enough of a prerequisite for the important recent advances in nonstationary time-series analysis, which are covered in the last two chapters of the book.

### **Designing a Course Out of the Book**

Several different courses can be designed based on the book.

- Assuming that the course meets twice for an hour and a half, eight weeks should be enough to cover core theory, which is Chapters 1–4 and 6 (excluding those

sections on specific economic applications), Chapter 7 (with proofs and examples skipped), and Chapter 8.

- A twelve-week semester course can cover, in addition to the core theory, Chapter 5 and the economic applications included in Chapters 1–6.
- A short (say, six-week) course specializing in GMM estimation in cross-section and panel data would cover Chapters 1–5 (excluding Sections 2.10–2.12 but including Section 2.2). Chapters 7 and 8 (excluding Section 8.7) would add the ML component to the course.
- A short time-series course covering recent developments with economic applications would consist of Chapter 1 (excluding Section 1.7), Chapter 2, parts of Chapters 6 and 7 (Sections 6.1–6.5, 6.8, and 7.1), and Chapters 9 and 10. The applications sections in Chapters 2, 6, 9, and 10 can be dropped if the course focuses on just theory.

### **Review Questions and Analytical Exercises**

The book includes a number of short questions for review at the end of each chapter, with lots of hints (and even answers). They can be used to check whether the reader actually understood the materials of the section. On the second reading, if not the first, readers should try to answer them before proceeding. Answers to selected review questions are available from the book's website, <http://pup.princeton.edu/titles/6946.html>.

There are several analytical exercises at the end of each chapter that ask the reader to prove results left unproved in the text or supplementary results that are useful for their own sake. Unless otherwise indicated, analytical exercises can be skipped without loss of continuity.

### **Empirical Exercises**

Each chapter usually has one big empirical exercise. It asks for a replication of the empirical results of the original article discussed in the applications section of the chapter and an estimation of the various extensions of the article's model using the estimation procedures presented in the chapter. The dataset for estimation, which usually is the same as the one used by the original article, can be downloaded from the book's website mentioned above.

To implement the estimation procedure on the dataset, readers should run a statistical package on a computer. There are many statistical packages that are widely used in econometrics. They include GAUSS ([www.aptech.com](http://www.aptech.com)), MATLAB ([www.mathworks.com](http://www.mathworks.com)), Eviews ([www.eviews.com](http://www.eviews.com)), LIMDEP ([www.limdep.com](http://www.limdep.com)),

RATS ([www.estima.com](http://www.estima.com)), SAS ([www.sas.com](http://www.sas.com)), Stata ([www.stata.com](http://www.stata.com)), and TSP ([www.tsp.com](http://www.tsp.com)). GAUSS and MATLAB are different from the rest in that they are matrix-based languages rather than a collection of procedures. Consider, for example, carrying out OLS with GAUSS or MATLAB. After loading the dataset into the computer's workspace, it takes several lines reflecting the matrix operations of OLS to calculate the OLS estimate and associated statistics (such as  $R^2$ ). With the other packages, which are procedure-driven and sometimes called "canned packages," those several lines can be replaced by a one-line command invoking the OLS procedure. For example, TSP's OLS command is `OLSQ`.

There are advantages and disadvantages with canned packages. Obviously, it takes far fewer lines to accomplish the same thing, so one can spend less time on programming. On the other hand, procedures in a canned package, which accept data and spit out point estimates and associated statistics, are essentially a black box. Sometimes it is not clear from the documentation of the package how certain statistics are calculated. Although those canned packages mentioned above regularly incorporate new developments in econometrics, the estimation procedure desired may not be currently supported by the package, in which case it will be necessary to write one's own procedures in GAUSS or MATLAB. But it may be a blessing in disguise; actually writing down the underlying matrix operations provides you with an excellent chance to understand the estimation procedure.

With only a few exceptions, all the calculations needed for the empirical exercises of the book can be carried out with any of the canned packages mentioned above. My recommendation, therefore, is for economics Ph.D. students who are planning to write an applied thesis using state-of-the-art estimation procedures or a theoretical thesis proposing new ones, to use GAUSS or MATLAB. Otherwise, students should use any of the canned packages mentioned above.

### **Mathematical Notation**

There is no single mathematical notation used by everyone in econometrics. The book's notation follows the most standard, if not universal, practice. Vectors are treated as column vectors and written in bold lowercase letters. Matrices are in bold uppercase letters. The transpose of the matrix  $\mathbf{A}$  is denoted by  $\mathbf{A}'$ . Scalar variables are (mostly lowercase) letters in italics.

### **Acknowledgments**

I acknowledge with gratitude help from the following individuals and institutions. Mark Watson, Dale Jorgenson, Bo Honore, Serena Ng, Masao Ogaki, and Jushan Bai were kind and courageous enough to use early versions of the book as the

textbook for their econometrics courses. Comments made by them and their students have been incorporated in this final version. Yuzo Honda read the manuscript and offered helpful suggestions. Naoto Kunitomo, Whitney Newey, Serena Ng, Pierre Perron, Jim Stock, Katsuto Tanaka, Mark Watson, Hal White, and Yoshihiro Yajima took time out to answer my questions and enquiries. Two graduate students at University of Tokyo, Mari Sakudo and Naoki Shimoi, read the entire manuscript to weed out typos. Their effort was underwritten by a grant-in-aid from the Zengin Foundation for Studies on Economics and Finance. Peter Dougherty, my editor at Princeton University Press, provided enthusiasm and just the right amount of pressure. Stephanie Hogue was a versatile enough  $\text{\LaTeX}$  expert to accommodate my formatting whims. Ellen Foos supervised production of the book. Last but not least, Jessica Helfand agreed, probably out of friendship, to do the cover design.

For more than five years all my free time went into writing this book. Now, having completed the book, I feel like someone who has just been released from prison. My research suffered, but hopefully the profession has not noticed.