

## PREFACE

### ECONOMETRIC ANALYSIS

*Econometric Analysis* provides a broad introduction to the field of econometrics. This field grows continually: a list of journals devoted at least in part, if not completely, to econometrics now includes *The Journal of Applied Econometrics*, *The Journal of Econometrics*, *The Econometrics Journal*, *Econometric Theory*, *Econometric Reviews*, *Journal of Business and Economic Statistics*, *Empirical Economics*, *Foundations and Trends in Econometrics*, *The Review of Economics and Statistics*, and *Econometrica*. Constructing a textbook style survey to introduce the topic at a graduate level has become increasingly ambitious. Nonetheless, I believe that one *can* successfully seek that objective in a single textbook. This text attempts to present, at an entry level, enough of the topics in econometrics that a student can comfortably move from here to practice or more advanced study in one or more specialized areas. The book is also intended as a bridge for students and analysts in the social sciences between an introduction to the field and the professional literature.

### NEW TO THIS EDITION

- Seventh*
- This 7<sup>th</sup> edition is a major revision of *Econometric Analysis*. Among the most obvious changes are
- Reorganization of the early material that is taught in the first semester course, including:
    - o All material on hypothesis testing and specification presented in a single chapter;
    - o New results on prediction;
    - o Greater and earlier emphasis on instrumental variables and endogeneity;
    - o Additional results on basic panel data models;
  - New applications and examples, with greater detail;
  - Greater emphasis on specific areas of application in the advanced material;
  - New material on simulation based methods, especially bootstrapping and Monte Carlo studies;
  - Several examples that explain interaction effects;
  - Specific recent applications including quantile regression;
  - New applications in discrete choice modeling;
  - New material on endogeneity and its implications for model structure.

### THE SEVENTH EDITION OF ECONOMETRIC ANALYSIS

The book has two objectives. The first is to introduce students to *applied econometrics*, including basic techniques in linear regression analysis and some of the rich variety of models that are used when the linear model proves inadequate or inappropriate. Modern software has made complicated modeling very easy to do, and an understanding of the underlying theory is also important. The second objective is to present students with sufficient *theoretical background* so that they will recognize new variants of the models learned about here as merely natural extensions that fit within a common body of principles. This book contains a substantial amount of theoretical material, such as that on GMM and maximum likelihood estimation and on asymptotic results for regression models.

This text is intended for a one-year graduate course for social scientists. Prerequisites should include calculus, mathematical statistics, and an introduction to econometrics at the level of, say, Gujarati's (2002) *Basic Econometrics*, Stock and Watson's (2006) *Introduction to Econometrics*, Kennedy's (2008) *Guide to Econometrics*, or Wooldridge's (2009) *Introductory Econometrics: A Modern Approach*. I assume, for example, that the reader has already learned about the basics of econometric methodology including the fundamental role of economic and statistical assumptions; the distinctions between cross section, time-series and panel data sets; and the essential ingredients of estimation, inference and prediction with the multiple linear regression model. Self-contained (for our purposes) summaries of the matrix algebra, mathematical statistics, and statistical theory used throughout the book are given in Appendices A through D. I rely heavily on matrix algebra throughout. This may be a bit daunting to some early on. But, matrix

algebra is an indispensable tool and I hope that the reader will come to agree that it is a means to an end, not an end in itself. With matrices, the unity of a variety of results will emerge without being obscured by a curtain of summation signs. All the matrix algebra needed in the text is presented in Appendix A. Appendix E and Chapter 15 contain a description of numerical methods that will be useful to practicing econometricians (and to us in the latter chapters of the book).

Contemporary computer software has made estimation of advanced nonlinear models as routine as least squares. I have included five chapters on estimation methods used in current research and seven chapters on applications in micro- and macroeconometrics. The nonlinear models used in these fields are now the staples of the applied econometrics literature. As a consequence, this book also contains a fair amount of material that will extend beyond many first courses in econometrics. Once again, I have included this in the hope of laying a foundation for study of the professional literature in these areas.

One overriding purpose has motivated all seven editions of this book. The vast majority of readers of this book will be users, not developers of econometrics. I believe that it is simply not sufficient to recite the theory of estimation, hypothesis testing, and econometric analysis. Although the often subtle theory is extremely important, the application is equally crucial. To that end, I have provided hundreds of numerical examples. My purpose in writing this work, and in my continuing efforts to update it, is to show readers how to *do* econometric analysis. I unabashedly accept the unflattering assessment of a correspondent who once likened this book to a "user's guide to econometrics."

## PLAN OF THE BOOK

The arrangement of the book is as follows:

Part I begins the formal development of econometrics with its fundamental pillar, the *linear multiple regression model*. Estimation and inference with the linear least squares estimator is analyzed in Chapters 2 through 6. The *nonlinear regression model* is introduced in Chapter 7 along with quantile, semi- and nonparametric regression, all as extensions of the familiar linear model. *Instrumental variables estimation* is developed in Chapter 8.

Part II presents three major extensions of the regression model. Chapter 9 presents the consequences of relaxing one of the main assumptions of the linear model, homoscedastic nonautocorrelated disturbances, to introduce the *generalized regression model*. The focus here is on heteroscedasticity; autocorrelation is mentioned, but a detailed treatment is deferred to Chapter 20 in the context of time series data. Chapter 10 introduces systems of regression equations, in principle as the approach to modeling simultaneously a set of random variables and, in practical terms, as an extension of the generalized linear regression model. Finally, panel data methods, primarily fixed and random effects models of heterogeneity, are presented in Chapter 11.

The second half of the book is devoted to topics that will extend the linear regression model in many directions. Beginning with Chapter 12, we proceed to the more involved methods of analysis that contemporary researchers use in analysis of "real-world" data. Chapters 12 to 16 in Part III present different estimation methodologies. Chapter 12 presents an overview by making the distinctions between *parametric*, *semiparametric* and *nonparametric methods*. The leading application of semiparametric estimation in the current literature is the *generalized method of moments (GMM) estimator* presented in Chapter 13. This technique provides the platform for much of modern econometrics. *Maximum likelihood estimation* is developed in Chapter 14. Monte Carlo and simulation-based methods such as bootstrapping, that have become a major component of current research are developed in Chapter 15. Finally, *Bayesian methods* are introduced in Chapter 16.

Parts IV and V develop two major subfields of econometric methods, *microeconometrics*, which is typically based on cross-section and panel data, and *macroeconometrics*, which is usually associated with analysis of time series data. In Part IV, Chapters 17 to 19 are concerned with models of discrete choice, censoring, truncation, sample selection, duration, treatment effects, and the analysis of counts of events. In Part V, Chapters 20 to 23, we consider time series models of serial correlation, lagged variables and nonstationary data — the usual substance of macroeconomic analysis.

## REVISIONS

sixth/  
sixth/  
is/

I have substantially rearranged the early part of the book to produce what I hope is a more natural sequence of topics for the graduate econometrics course. Chapter 4 is now devoted entirely to point and interval estimation, including prediction and forecasting. Finite sample, then asymptotic properties of least squares are developed in detail. All of the material on hypothesis testing and specification search is moved into Chapter 5, rather than fragmented over several chapters as in the 6<sup>th</sup> edition. I have also brought the material on instrumental variables much farther forward in the text, from after the development of the generalized regression model in the 6<sup>th</sup> edition to Chapter 8 in this one, immediately after full development of the linear regression model. This accords with the greater emphasis on this method in recent applications. A very large number of other rearrangements of the material will also be evident. Chapter 7 now contains a range of advanced extensions of the linear regression model, including nonlinear, quantile, partially linear and nonparametric regression. This is also a point at which the differences between parametric, semiparametric and nonparametric methods can be examined. One conspicuous modification is the excision of the long chapter on linear simultaneous equations models. Some of the material from this chapter appears elsewhere. Two stage least squares now appears with instrumental variables estimation, while the analysis of properties of dynamic models will be presented in Chapter 21 where we examine models with lagged variables. Remaining parts of this chapter that are of lesser importance in recent treatments, such as rank and order conditions for identification of linear models and 3SLS and FIML estimation, have been deleted or greatly reduced and placed in context elsewhere in the text. The material on discrete choice models has been rearranged to orient the topics to the behavioral foundations. Chapter 17 now broadly introduces discrete choice, random utility models, then builds on variants of the binary choice model. The analysis is continued in Chapter 18 with unordered, then ordered choice models and, finally models for counts. The last chapter of the section studies models for continuous variables in the contexts of particular data generating mechanisms and behavioral contexts.

I have added new material and some different examples and applications at numerous points. Topics that have been expanded or given greater emphasis include treatment effects, bootstrapping, simulation based estimation, robust estimation, missing and faulty data, and a variety of different new methods of discrete choice analysis in microeconometrics. I have also added or expanded material on techniques recently of interest, such as quantile regression and stochastic frontier models.

I note a few specific highlights of the revision: In general terms, I have increased the focus on robust methods a bit. I have placed discussions of specification tests at several points, consistent with the trend in the literature to examine more closely the fragility of heavily parametric models. A few of the specific new applications are as follows:

- Chapter 15 on simulation based estimation has been considerably expanded. It now includes substantially more material on bootstrapping standard errors and confidence intervals. The Krinsky and Robb (1986) approach to asymptotic inference has been placed here as well.
- A great deal of attention has been focused in recent papers on how to understand interaction effects in nonlinear models. Chapter 7 contains a lengthy application of interaction effects in a nonlinear (exponential) regression model. The issue is revisited in Chapter 17.
- As an exercise that will challenge the student's facility with asymptotic distribution theory, I have added a detailed proof of the Murphy and Topel (2002) result for two step estimation in Chapter 14.
- Sources and treatment of endogeneity appear at various points, for example an application of inverse probability weighting to deal with attrition in Chapter 17.

seventh

The 7<sup>th</sup> edition is a major revision of *Econometric Analysis* both in terms of organization of the material and in terms of new ideas and treatments. I hope that readers will find the changes helpful.

## SOFTWARE AND DATA

There are many computer programs that are widely used for the computations described in this book. All were written by econometricians or statisticians, and in general, all are regularly updated to incorporate new developments in applied econometrics. A sampling of the most widely used packages and Internet home pages where you can find information about them are:

<u>EVIEWS</u>	<u>www.eviews.com</u>	(QMS, Irvine, CA),
<u>Gauss</u>	<u>www.aptech.com</u>	(Aptech Systems, Kent, WA),
<u>LIMDEP</u>	<u>www.limdep.com</u>	(Econometric Software, Plainview, NY),
<u>MATLAB</u>	<u>www.mathworks.com</u>	(Mathworks, Natick, MA),
<u>NLOGIT</u>	<u>www.nlogit.com</u>	(Econometric Software, Plainview, NY),
<u>R</u>	<u>www.r-project.org/</u>	(The R Project for Statistical Computing),
<u>RATS</u>	<u>www.estima.com</u>	(Estima, Evanston, IL),
<u>SAS</u>	<u>www.sas.com</u>	(SAS, Cary, NC),
<u>Shazam</u>	<u>econometrics.com</u>	(Northwest Econometrics Ltd, Gibsons, Canada),
<u>Stata</u>	<u>www.stata.com</u>	(Stata, College Station, TX),
<u>TSP</u>	<u>www.tspintl.com</u>	(TSP International, Stanford, CA).

A more extensive list of computer software used for econometric analysis can be found at the resource website, <http://www.oswego.edu/~economic/econsoftware.htm>.

With only a few exceptions, the computations described in this book can be carried out with any of the packages listed. *NLOGIT* was used for the computations in the applications to follow. This text contains no instruction on using any particular program or language. (The author's website for the text does provide some code and data for replication of the numerical examples.) Many authors have produced *RATS*, *LIMDEP/NLOGIT*, *EVIEWS*, *SAS* or *Stata* code for some of our applications, including, in a few cases, in the documentations for their computer programs. There are also quite a few volumes now specifically devoted to econometrics associated with particular packages, such as Cameron and Trivedi's (2009) companion to their treatise on microeconometrics..

The data sets used in the examples are also available on the Web site for the text, <http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm>. Throughout the text, these data sets are referred to "Table Fn.m," for example Table F4.1. The "F" refers to Appendix F at the back of the text, which contains descriptions of the data sets. The actual data are posted in generic ASCII and portable formats on the Web site with the other supplementary materials for the text. There are now thousands of interesting Web sites containing software, data sets, papers, and commentary on econometrics. It would be hopeless to attempt any kind of a survey here. One code/data site that is particularly agreeably structured and well targeted for readers of this book is the data archive for the *Journal of Applied Econometrics*. They have archived all the nonconfidential data sets used in their publications since 1988 (with some gaps before 1995). This useful site can be found at <http://qed.econ.queensu.ca/jae/>. Several of the examples in the text use the *JAE* data sets. Where we have done so, we direct the reader to the *JAE*'s website, rather than our own, for replication. Other journals have begun to ask their authors to provide code and data to encourage replication. Another vast, easy to navigate site for aggregate data on the U.S. economy is [www.econmagic.com](http://www.econmagic.com).

## ACKNOWLEDGMENTS

It is a pleasure to express my appreciation to those who have influenced this work. I remain grateful to Arthur Goldberger (dec.), Arnold Zellner (dec.), Dennis Aigner, Bill Becker, and Laurits Christensen for their encouragement and guidance. After seven editions of this book, the number of individuals who have significantly improved it through their comments, criticisms, and encouragement has become far too large for me to thank all of them individually. I am grateful for their help and I hope that all of them see their contribution to this edition. I would like to acknowledge the many reviewers of my work whose careful reading has vastly improved the book through this edition: Scott Atkinson, University of Georgia; Badi Baltagi, Syracuse University; Neal Beck, New York University; William E. Becker (Ret.), Indiana University; Eric J. Belasko, Texas Tech University; Anil Bera, University of Illinois; John Burkett, University of Rhode Island; Leonard Carlson, Emory University; Frank Chaloupka, University of Illinois at Chicago; Chris Cornwell, University of Georgia; Craig Depken II, University of Texas at Arlington; Frank Diebold, University of Pennsylvania; Edward Dwyer, Clemson University; Michael Ellis, Wesleyan University; Martin Evans, Georgetown University; Vahagn Galstyan, Trinity College Dublin; Paul Glewwe, University of Minnesota; Ed Greenberg, Washington University at St. Louis; Miguel Herce, University of North Carolina; Joseph Hilbe, Arizona State University; Dr. Uwe Jensen, Christian-Albrecht University; K. Rao Kadiyala, Purdue University; William Lott, University of Connecticut; Thomas L. Marsh, Washington State University; Edward Mathis, Villanova University; Mary McGarvey, University of Nebraska-Lincoln; Ed Melnick, New York University; Thad Mirer, State University of New York at Albany; Cyril Pasche, University of Geneva; Paul Ruud, University of California at Berkeley; Sherrie Rhine, Federal Deposit Insurance Corp.; Terry G. Seaks (Ret.), University of North Carolina at Greensboro; Donald Snyder, California State University at Los Angeles; Steven Stern, University of Virginia; Houston Stokes, University of Illinois at Chicago; Dmitrios Thomakos, Columbia University; Paul Wachtel, New York University; Mary Beth Walker, Georgia State University; Mark Watson, Harvard University; and Kenneth West, University of Wisconsin. My numerous discussions with Bruce McCullough of Drexel University have improved Appendix E and at the same time increased my appreciation for numerical analysis. I am especially grateful to Jan Kiviet of the University of Amsterdam, who subjected my third edition to a microscopic examination and provided literally scores of suggestions, virtually all of which appear herein. Professor Pedro Bacao, University of Coimbra, Portugal and Mark Strahan of Sand Hill Econometrics did likewise with the 6<sup>th</sup> edition.

I've had great support and encouragement over the years from many people close to me, especially my family, and many not so close. None has been more gratifying than the mail I've received from readers from the world over who have shared my enthusiasm for this exciting field and for this work that has taught them and me econometrics since the first edition in 1990.

Finally, I would also like to thank the many people at Prentice Hall who have put this book together with me: Adrienne D'Ambrosio, Jill Kolongowski, Carla Thompson, Nancy Fenton, Alison Eusden, Joe Vetere, Martha Wetherill, and the composition team at <insert name of composition team>.

William H. Greene  
August, 2010

each /

sixth /

AV:  
Provide  
missing  
names