

高等学校经济类双语教学推荐教材

PEARSON



ECONOMICS
CLASSICS

经济学经典教材·核心课系列

Economics Classics

计量经济分析

Econometric
Analysis

经济分析 上册

(英文版·第七版)

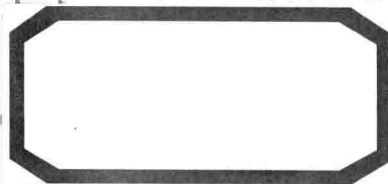
(Seventh Edition)

威廉·H·格林 (William H. Greene) 著

中国人民大学出版社



高等学校经济类双语教学推荐教材



ECONOMICS
CLASSICS

经济学经典教材·核心课系列

Classics

计量经济分析 上册

Econometric Analysis

(英文版·第七版)



威廉·H·格林

中国人民大学出版社

· 北京 ·

出版说明

入世十年，我国已完全融入到经济全球化的浪潮中。党的十六大确立了“引进来，走出去”的发展战略，使得“国际化”复合型人才的需求不断增加。这就对我国一般本科院校多年来所采取的单一语言（母语）教学提出了严峻挑战，经济类专业双语教学改革迫在眉睫。

为配合高校经济类专业双语教学改革，中国人民大学出版社携手培生、麦格劳-希尔、圣智等众多国际知名出版公司，倾情打造了该套“经济类双语系列教材”，本套教材包括：经济管理类专业开设的核心课程、经济学专业开设的主干课程以及财政金融专业和国际贸易专业的主要课程。所选教材均为国外最优秀的本科层次经济类教材。

我们在组织、引进和出版该系列教材的过程中，严把质量关。聘请国内著名经济学家、学者以及一线授课教师审核国外原版教材，广泛听取意见，努力做到把国外真正高水平的适合国内实际教学需求的优秀教材引进来，供国内广大师生参考、研究和学习。

本系列教材主要有以下特点：

第一，教材体系设计完整。本系列教材全部为国外知名出版公司的优秀教材，涵盖了经济类专业的所有主要课程。

第二，保持英文原版教材特色。本系列教材依据国内实际教学需要以及广泛的适应性，部分对原版教材进行了全文影印，部分在保持原版教材体系结构和内容特色的基础上进行了适当删减。

第三，内容紧扣学科前沿。本系列教材在原著选择上紧扣国外教学前沿，基本上都是国外最流行教材的最新版本。

第四，篇幅合理、价格适中。本系列教材一方面在内容和篇幅上很好地适应了国内双语教学的实际需要，另一方面，低定价策略又避免了国外原版图书高额的购买费用。

第五，提供强大的教学支持。依托国外知名出版公司的资源，本系列教材为教师提供丰富的配套教辅资源，如教师手册、PPT课堂演示文稿、试题库等，并配套有内容丰富的网络资源，使教学更为便利。

本系列教材既适合高等院校经济类专业的本科教学使用，也适合从事经济类工作和研究的广大从业者阅读和学习。我们在选书、改编过程中虽然全面听取了专家、学者和教师的意见，努力做到满足广大读者的需求，但由于各教材的作者所处的政治、经济和文化背景不同，书中内容仍可能有不妥之处，我们真诚希望广大读者提出宝贵意见和建议，以便我们在以后的版本中不断改进和完善。

中国人民大学出版社

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

This seventh 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
 - All material on hypothesis testing and specification presented in a single chapter
 - New results on prediction
 - Greater and earlier emphasis on instrumental variables and endogeneity
 - 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, maximum likelihood estimation, and 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 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 later 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 five 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 are 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 and 21, we consider two topics in time-series analysis, models of serial correlation and regression models for nonstationary data—the usual substance of macroeconomic analysis.

REVISIONS

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 sixth 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 sixth 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. 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 and random utility models, and 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.

The seventh 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

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

A more extensive list of computer software used for econometric analysis can be found at the resource Web site, <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. This text contains no instruction on using any particular program or language. (The author's Web site 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, the documentation 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 Web site, 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.economagic.com.

目 录

第一部分 线性回归模型

第 1 章	计量经济学	1
第 2 章	线性回归模型	9
第 3 章	最小二乘法	22
第 4 章	最小二乘估计	43
第 5 章	假设检验和模型选择	90
第 6 章	函数形式与结构变化	124
第 7 章	非线性、半参数和非参数回归模型	151
第 8 章	内生性和工具变量估计	183

第二部分 广义回归模型与联立方程模型

第 9 章	广义回归模型与异方差	214
第 10 章	方程组	241
第 11 章	面板数据模型	286

第三部分 估计方法

第 12 章	计量经济学的估计框架	360
第 13 章	最小距离估计与广义矩估计法	379
第 14 章	极大似然估计	424
第 15 章	基于模拟的估计与统计推断及随机参数模型	502
第 16 章	贝叶斯估计与统计推断	545

第四部分 横截面、面板数据及微观计量学

第 17 章	离散选择	566
第 18 章	离散选择与事件研究	631
第 19 章	受限因变量——截断、设限与样本选择	692

第五部分 时间序列与宏观计量学

第 20 章	序列相关	750
第 21 章	非平稳数据	782

第六部分 附录

附录 A	矩阵代数	808
附录 B	概率与分布理论	843
附录 C	估计与统计推断	870
附录 D	大样本分布理论	886
附录 E	计算与优化	905
附录 F	实际应用中的数据	921
参考文献		927

CONTENTS



PART I The Linear Regression Model

CHAPTER 1 Econometrics 1

- 1.1 Introduction 1
- 1.2 The Paradigm of Econometrics 1
- 1.3 The Practice of Econometrics 3
- 1.4 Econometric Modeling 3
- 1.5 Plan of the Book 6
- 1.6 Preliminaries 7

CHAPTER 2 The Linear Regression Model 9

- 2.1 Introduction 9
- 2.2 The Linear Regression Model 10
- 2.3 Assumptions of the Linear Regression Model 12
- 2.4 Summary and Conclusions 20

CHAPTER 3 Least Squares 22

- 3.1 Introduction 22
- 3.2 Least Squares Regression 22
- 3.3 Partitioned Regression and Partial Regression 27
- 3.4 Partial Regression and Partial Correlation Coefficients 30
- 3.5 Goodness of Fit and the Analysis of Variance 32
- 3.6 Linearly Transformed Regression 38
- 3.7 Summary and Conclusions 39

CHAPTER 4 The Least Squares Estimator 43

- 4.1 Introduction 43
- 4.2 Motivating Least Squares 44
- 4.3 Finite Sample Properties of Least Squares 45
- 4.4 Large Sample Properties of the Least Squares Estimator 53
- 4.5 Interval Estimation 63
- 4.6 Prediction and Forecasting 67
- 4.7 Data Problems 74
- 4.8 Summary and Conclusions 85

CHAPTER 5 Hypothesis Tests and Model Selection 90

- 5.1 Introduction 90
- 5.2 Hypothesis Testing Methodology 90
- 5.3 Two Approaches to Testing Hypotheses 94
- 5.4 Wald Tests Based on the Distance Measure 95
- 5.5 Testing Restrictions Using the Fit of the Regression 101
- 5.6 Nonnormal Disturbances and Large-Sample Tests 106
- 5.7 Testing Nonlinear Restrictions 109
- 5.8 Choosing between Nonnested Models 111
- 5.9 A Specification Test 114
- 5.10 Model Building—A General to Simple Strategy 115
- 5.11 Summary and Conclusions 119

CHAPTER 6 Functional Form and Structural Change 124

- 6.1 Introduction 124
- 6.2 Using Binary Variables 124
- 6.3 Nonlinearity in the Variables 131
- 6.4 Modeling and Testing for a Structural Break 140
- 6.5 Summary and Conclusions 146

CHAPTER 7 Nonlinear, Semiparametric, and Nonparametric Regression Models 151

- 7.1 Introduction 151
- 7.2 Nonlinear Regression Models 152
- 7.3 Median and Quantile Regression 169
- 7.4 Partially Linear Regression 175
- 7.5 Nonparametric Regression 176
- 7.6 Summary and Conclusions 179

CHAPTER 8 Endogeneity and Instrumental Variable Estimation 183

- 8.1 Introduction 183
- 8.2 Assumptions of the Extended Model 186
- 8.3 Estimation 187
- 8.4 Two Specification Tests 195
- 8.5 Measurement Error 200
- 8.6 Nonlinear Instrumental Variables Estimation 205
- 8.7 Weak Instruments 208
- 8.8 Natural Experiments and the Search for Causal Effects 209
- 8.9 Summary and Conclusions 212

PART II Generalized Regression Model and Equation Systems

CHAPTER 9 The Generalized Regression Model and Heteroscedasticity 214

- 9.1 Introduction 214
- 9.2 Inefficient Estimation by Least Squares and Instrumental Variables 215

- 9.3 Efficient Estimation by Generalized Least Squares 220
- 9.4 Heteroscedasticity and Weighted Least Squares 223
- 9.5 Testing for Heteroscedasticity 228
- 9.6 Weighted Least Squares 231
- 9.7 Applications 233
- 9.8 Summary and Conclusions 237

CHAPTER 10 Systems of Equations 241

- 10.1 Introduction 241
- 10.2 The Seemingly Unrelated Regressions Model 242
- 10.3 Seemingly Unrelated Generalized Regression Models 253
- 10.4 Nonlinear Systems of Equations 253
- 10.5 Systems of Demand Equations: Singular Systems 255
- 10.6 Simultaneous Equations Models 261
- 10.7 Summary and Conclusions 279

CHAPTER 11 Models for Panel Data 286

- 11.1 Introduction 286
- 11.2 Panel Data Models 286
- 11.3 The Pooled Regression Model 291
- 11.4 The Fixed Effects Model 299
- 11.5 Random Effects 309
- 11.6 Nonspherical Disturbances and Robust Covariance Estimation 321
- 11.7 Spatial Autocorrelation 324
- 11.8 Endogeneity 328
- 11.9 Nonlinear Regression with Panel Data 342
- 11.10 Systems of Equations 346
- 11.11 Parameter Heterogeneity 347
- 11.12 Summary and Conclusions 355

PART III Estimation Methodology

CHAPTER 12 Estimation Frameworks in Econometrics 360

- 12.1 Introduction 360
- 12.2 Parametric Estimation and Inference 361
- 12.3 Semiparametric Estimation 366
- 12.4 Nonparametric Estimation 370
- 12.5 Properties of Estimators 373
- 12.6 Summary and Conclusions 378

CHAPTER 13 Minimum Distance Estimation and the Generalized Method of Moments 379

- 13.1 Introduction 379
- 13.2 Consistent Estimation: The Method of Moments 380

- 13.3 Minimum Distance Estimation 386
- 13.4 The Generalized Method of Moments (GMM) Estimator 390
- 13.5 Testing Hypotheses in the GMM Framework 399
- 13.6 GMM Estimation of Econometric Models 401
- 13.7 Summary and Conclusions 422

CHAPTER 14 Maximum Likelihood Estimation 424

- 14.1 Introduction 424
- 14.2 The Likelihood Function and Identification of the Parameters 424
- 14.3 Efficient Estimation: The Principle of Maximum Likelihood 426
- 14.4 Properties of Maximum Likelihood Estimators 428
- 14.5 Conditional Likelihoods, Econometric Models, and the GMM Estimator 436
- 14.6 Hypothesis and Specification Tests and Fit Measures 437
- 14.7 Two-Step Maximum Likelihood Estimation 446
- 14.8 Pseudo-Maximum Likelihood Estimation and Robust Asymptotic Covariance Matrices 451
- 14.9 Applications of Maximum Likelihood Estimation 456
- 14.10 Latent Class and Finite Mixture Models 490
- 14.11 Summary and Conclusions 498

CHAPTER 15 Simulation-Based Estimation and Inference and Random Parameter Models 502

- 15.1 Introduction 502
- 15.2 Random Number Generation 504
- 15.3 Simulation-Based Statistical Inference: The Method of Krinsky and Robb 507
- 15.4 Bootstrapping Standard Errors and Confidence Intervals 509
- 15.5 Monte Carlo Studies 512
- 15.6 Simulation-Based Estimation 517
- 15.7 A Random Parameters Linear Regression Model 527
- 15.8 Hierarchical Linear Models 532
- 15.9 Nonlinear Random Parameter Models 534
- 15.10 Individual Parameter Estimates 534
- 15.11 Mixed Models and Latent Class Models 541
- 15.12 Summary and Conclusions 543

CHAPTER 16 Bayesian Estimation and Inference 545

- 16.1 Introduction 545
- 16.2 Bayes Theorem and the Posterior Density 546
- 16.3 Bayesian Analysis of the Classical Regression Model 547
- 16.4 Bayesian Inference 552
- 16.5 Posterior Distributions and the Gibbs Sampler 556
- 16.6 Application: Binomial Probit Model 558
- 16.7 Panel Data Application: Individual Effects Models 560
- 16.8 Hierarchical Bayes Estimation of a Random Parameters Model 562
- 16.9 Summary and Conclusions 564

1

ECONOMETRICS



1.1 INTRODUCTION

This book will present an introductory survey of econometrics. We will discuss the fundamental ideas that define the methodology and examine a large number of specific models, tools and methods that econometricians use in analyzing data. This chapter will introduce the central ideas that are the paradigm of econometrics. Section 1.2 defines the field and notes the role that theory plays in motivating econometric practice. Section 1.3 discusses the types of applications that are the focus of econometric analyses. The process of econometric modeling is presented in Section 1.4 with a classic application, Keynes's consumption function. A broad outline of the book is presented in Section 1.5. Section 1.6 notes some specific aspects of the presentation, including the use of numerical examples and the mathematical notation that will be used throughout the book.

1.2 THE PARADIGM OF ECONOMETRICS

In the first issue of *Econometrica*, Ragnar Frisch (1933) said of the Econometric Society that

its main object shall be to promote studies that aim at a unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems and that are penetrated by constructive and rigorous thinking similar to that which has come to dominate the natural sciences. But there are several aspects of the quantitative approach to economics, and no single one of these aspects taken by itself, should be confounded with econometrics. Thus, econometrics is by no means the same as economic statistics. Nor is it identical with what we call general economic theory, although a considerable portion of this theory has a definitely quantitative character. Nor should econometrics be taken as synonymous [*sic*] with the application of mathematics to economics. Experience has shown that each of these three viewpoints, that of statistics, economic theory, and mathematics, is a necessary, but not by itself a sufficient, condition for a real understanding of the quantitative relations in modern economic life. It is the *unification* of all three that is powerful. And it is this unification that constitutes econometrics.

The Society responded to an unprecedented accumulation of statistical information. They saw a need to establish a body of principles that could organize what would otherwise become a bewildering mass of data. Neither the pillars nor the objectives of econometrics have changed in the years since this editorial appeared. Econometrics concerns itself with the application of mathematical statistics and the tools of statistical inference to the empirical measurement of relationships postulated by an underlying theory.

The crucial role that econometrics plays in economics has grown over time. The Nobel Prize in Economic Sciences has recognized this contribution with numerous awards to econometricians, including the first which was given to (the same) Ragnar Frisch in 1969, Lawrence Klein in 1980, Trygve Haavelmo in 1989, James Heckman and Daniel McFadden in 2000, and Robert Engle and Clive Granger in 2003. The 2000 prize was noteworthy in that it celebrated the work of two scientists whose research was devoted to the marriage of behavioral theory and econometric modeling.

Example 1.1 Behavioral Models and the Nobel Laureates

The pioneering work by both James Heckman and Dan McFadden rests firmly on a theoretical foundation of utility maximization.

For Heckman's, we begin with the standard theory of household utility maximization over consumption and leisure. The textbook model of utility maximization produces a demand for leisure time that translates into a supply function of labor. When home production (work in the home as opposed to the outside, formal labor market) is considered in the calculus, then desired "hours" of (formal) labor can be negative. An important conditioning variable is the "reservation" wage—the wage rate that will induce formal labor market participation. On the demand side of the labor market, we have firms that offer market wages that respond to such attributes as age, education, and experience. What can we learn about labor supply behavior based on observed market wages, these attributes and observed hours in the formal market? Less than it might seem, intuitively because our observed data omit half the market—the data on formal labor market activity are not randomly drawn from the whole population.

Heckman's observations about this implicit truncation of the distribution of hours or wages revolutionized the analysis of labor markets. Parallel interpretations have since guided analyses in every area of the social sciences. The analysis of policy interventions such as education initiatives, job training and employment policies, health insurance programs, market creation, financial regulation and a host of others is heavily influenced by Heckman's pioneering idea that when participation is part of the behavior being studied, the analyst must be cognizant of the impact of common influences in both the presence of the intervention and the outcome. We will visit the literature on sample selection and treatment/program evaluation in Chapter 18.

Textbook presentations of the theories of demand for goods that produce utility, since they deal in continuous variables, are conspicuously silent on the kinds of discrete choices that consumers make every day—what brand of product to choose, whether to buy a large commodity such as a car or a refrigerator, how to travel to work, whether to rent or buy a home, where to live, what candidate to vote for, and so on. Nonetheless, a model of "random utility" defined over the alternatives available to the consumer provides a theoretically sound platform for studying such choices. Important variables include, as always, income and relative prices. What can we learn about underlying preference structures from the discrete choices that consumers make? What must be assumed about these preferences to allow this kind of inference? What kinds of statistical models will allow us to draw inferences about preferences? McFadden's work on how commuters choose to travel to work, and on the underlying theory appropriate to this kind of modeling, has guided empirical research in discrete consumer choices for several decades. We will examine McFadden's models of discrete choice in Chapter 18.

The connection between underlying behavioral models and the modern practice of econometrics is increasingly strong. A useful distinction is made between *microeconometrics* and *macroeconometrics*. The former is characterized by its analysis of cross section and panel data and by its focus on individual consumers, firms, and micro-level decision makers. Practitioners rely heavily on the theoretical tools of microeconomics including utility maximization, profit maximization, and market equilibrium. The analyses are directed at subtle, difficult questions that often require intricate formulations. A few applications are as follows:

- What are the likely effects on labor supply behavior of proposed negative income taxes? [Ashenfelter and Heckman (1974).]
- Does attending an elite college bring an expected payoff in expected lifetime income sufficient to justify the higher tuition? [Kreuger and Dale (1999) and Kreuger (2000).]
- Does a voluntary training program produce tangible benefits? Can these benefits be accurately measured? [Angrist (2001).]
- Do smaller class sizes bring real benefits in student performance? [Hanushek (1999), Hoxby (2000), Angrist and Lavy (1999).]
- Does the presence of health insurance induce individuals to make heavier use of the health care system—is moral hazard a measurable problem? [Riphahn et al. (2003).]

Macroeconometrics is involved in the analysis of time-series data, usually of broad aggregates such as price levels, the money supply, exchange rates, output, investment, economic growth and so on. The boundaries are not sharp. For example, an application that we will examine in this text concerns spending patterns of municipalities, which

rests somewhere between the two fields. The very large field of financial econometrics is concerned with long time-series data and occasionally vast panel data sets, but with a sharply focused orientation toward models of individual behavior. The analysis of market returns and exchange rate behavior is neither exclusively macro- nor microeconomic. (We will not be spending any time in this book on financial econometrics. For those with an interest in this field, I would recommend the celebrated work by Campbell, Lo, and Mackinlay (1997) or, for a more time-series-oriented approach, Tsay (2005).) Macroeconomic model builders rely on the interactions between economic agents and policy makers. For examples:

- Does a monetary policy regime that is strongly oriented toward controlling inflation impose a real cost in terms of lost output on the U.S. economy? [Cecchetti and Rich (2001).]
- Did 2001's largest federal tax cut in U.S. history contribute to or dampen the concurrent recession? Or was it irrelevant?

Each of these analyses would depart from a formal model of the process underlying the observed data.

1.3 THE PRACTICE OF ECONOMETRICS

We can make another useful distinction between *theoretical econometrics* and *applied econometrics*. Theorists develop new techniques for estimation and hypothesis testing and analyze the consequences of applying particular methods when the assumptions that justify those methods are not met. Applied econometricians are the users of these techniques and the analysts of data ("real world" and simulated). The distinction is far from sharp; practitioners routinely develop new analytical tools for the purposes of the study that they are involved in. This book contains a large amount of econometric theory, but it is directed toward applied econometrics. I have attempted to survey techniques, admittedly some quite elaborate and intricate, that have seen wide use "in the field."

Applied econometric methods will be used for estimation of important quantities, analysis of economic outcomes such as policy changes, markets or individual behavior, testing theories, and for forecasting. The last of these is an art and science in itself that is the subject of a vast library of sources. Although we will briefly discuss some aspects of forecasting, our interest in this text will be on estimation and analysis of models. The presentation, where there is a distinction to be made, will contain a blend of microeconomic and macroeconomic techniques and applications. It is also necessary to distinguish between *time-series analysis* (which is not our focus) and methods that primarily use time-series data. The former is, like forecasting, a growth industry served by its own literature in many fields. While we will employ some of the techniques of time-series analysis, we will spend relatively little time developing first principles.

1.4 ECONOMETRIC MODELING

Econometric analysis usually begins with a statement of a theoretical proposition. Consider, for example, a classic application by one of Frisch's contemporaries:

Example 1.2 Keynes's Consumption Function

From Keynes's (1936) *General Theory of Employment, Interest and Money*:

We shall therefore define what we shall call the propensity to consume as the functional relationship f between X , a given level of income, and C , the expenditure on consumption out of the level of income, so that $C = f(X)$.

The amount that the community spends on consumption depends (i) partly on the amount of its income, (ii) partly on other objective attendant circumstances, and (iii) partly on the subjective needs and the psychological propensities and habits of the individuals composing it. The fundamental psychological law upon which we are entitled to depend with great confidence, both a priori from our knowledge of human