

## 1

## INTRODUCTION



## 1.1 ECONOMETRICS

In the first issue of *Econometrica*, the Econometric Society stated 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.

Frisch (1933) and his 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 is the field of economics that concerns itself with the application of mathematical statistics and the tools of statistical inference to the empirical measurement of relationships postulated by economic theory.

## 1.2 ECONOMETRIC MODELING

Econometric analysis will usually begin with a statement of a theoretical proposition. Consider, for example, a canonical application:

**Example 1.1 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

## 2 CHAPTER 1 ♦ Introduction

(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 nature and from the detailed facts of experience, is that men are disposed, as a rule and on the average, to increase their consumption as their income increases, but not by as much as the increase in their income.<sup>1</sup> That is,  $\dots dC/dX$  is positive and less than unity.

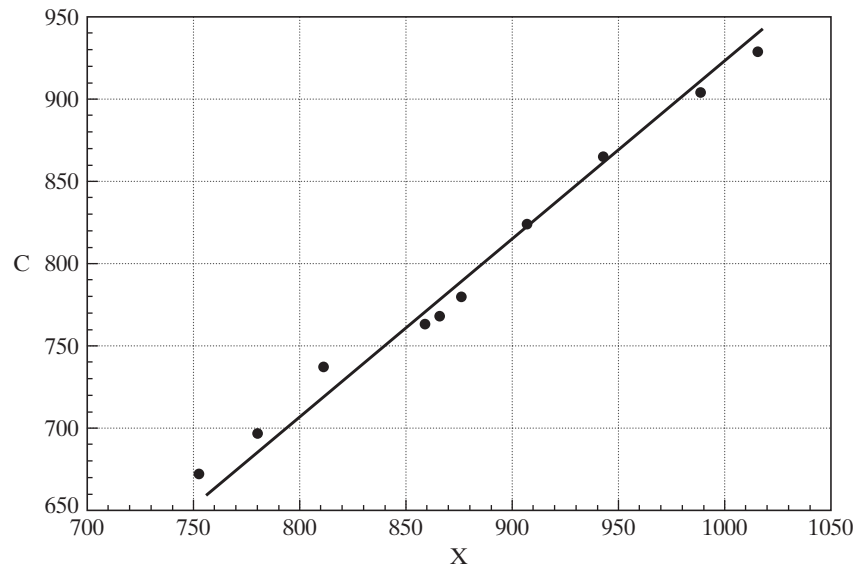
But, apart from short period changes in the level of income, it is also obvious that a higher absolute level of income will tend as a rule to widen the gap between income and consumption. . . . These reasons will lead, as a rule, to a greater proportion of income being saved as real income increases.

The theory asserts a relationship between consumption and income,  $C = f(X)$ , and claims in the third paragraph that the marginal propensity to consume (MPC),  $dC/dX$ , is between 0 and 1. The final paragraph asserts that the average propensity to consume (APC),  $C/X$ , falls as income rises, or  $d(C/X)/dX = (MPC - APC)/X < 0$ . It follows that  $MPC < APC$ . The most common formulation of the consumption function is a linear relationship,  $C = \alpha + \beta X$ , that satisfies Keynes's "laws" if  $\beta$  lies between zero and one and if  $\alpha$  is greater than zero.

These theoretical propositions provide the basis for an econometric study. Given an appropriate data set, we could investigate whether the theory appears to be consistent with the observed "facts." For example, we could see whether the linear specification appears to be a satisfactory description of the relationship between consumption and income, and, if so, whether  $\alpha$  is positive and  $\beta$  is between zero and one. Some issues that might be studied are (1) whether this relationship is stable through time or whether the parameters of the relationship change from one generation to the next (a change in the average propensity to save,  $1 - APC$ , might represent a fundamental change in the behavior of consumers in the economy); (2) whether there are systematic differences in the relationship across different countries, and, if so, what explains these differences; and (3) whether there are other factors that would improve the ability of the model to explain the relationship between consumption and income. For example, Figure 1.1 presents aggregate consumption and personal income in constant dollars for the U.S. for the 10 years of 1970–1979. (See Appendix Table F1.1.) Apparently, at least superficially, the data (the facts) are consistent with the theory. The relationship appears to be linear, albeit only approximately, the intercept of a line that lies close to most of the points is positive and the slope is less than one, although not by much.

Economic theories such as Keynes's are typically crisp and unambiguous. Models of demand, production, and aggregate consumption all specify precise, *deterministic* relationships. Dependent and independent variables are identified, a functional form is specified, and in most cases, at least a qualitative statement is made about the directions of effects that occur when independent variables in the model change. Of course, the model is only a simplification of reality. It will include the salient features of the relationship of interest, but will leave unaccounted for influences that might well be present but are regarded as unimportant. No model could hope to encompass the myriad essentially random aspects of economic life. It is thus also necessary to incorporate stochastic elements. As a consequence, observations on a dependent variable will display variation attributable not only to differences in variables that are explicitly accounted for, but also to the randomness of human behavior and the interaction of countless minor influences that are not. It is understood that the introduction of a random "disturbance" into a deterministic model is not intended merely to paper over its inadequacies. It is

<sup>1</sup>Modern economists are rarely this confident about their theories. More contemporary applications generally begin from first principles and behavioral axioms, rather than simple observation.



**FIGURE 1.1** Consumption Data, 1970–1979.

essential to examine the results of the study, in a sort of postmortem, to ensure that the allegedly random, unexplained factor is truly unexplainable. If it is not, the model is, in fact, inadequate. The stochastic element endows the model with its statistical properties. Observations on the variable(s) under study are thus taken to be the outcomes of a random process. With a sufficiently detailed stochastic structure and adequate data, the analysis will become a matter of deducing the properties of a probability distribution. The tools and methods of mathematical statistics will provide the operating principles.

A model (or theory) can never truly be confirmed unless it is made so broad as to include every possibility. But it may be subjected to ever more rigorous scrutiny and, in the face of contradictory evidence, refuted. A deterministic theory will be invalidated by a single contradictory observation. The introduction of stochastic elements into the model changes it from an exact statement to a probabilistic description about expected outcomes and carries with it an important implication. Only a preponderance of contradictory evidence can convincingly invalidate the probabilistic model, and what constitutes a “preponderance of evidence” is a matter of interpretation. Thus, the probabilistic model is less precise but at the same time, more robust.<sup>2</sup>

The process of econometric analysis departs from the specification of a theoretical relationship. We initially proceed on the optimistic assumption that we can obtain precise measurements on all the variables in a correctly specified model. If the ideal conditions are met at every step, the subsequent analysis will probably be routine. Unfortunately, they rarely are. Some of the difficulties one can expect to encounter are the following:

<sup>2</sup>See Keuzenkamp and Magnus (1995) for a lengthy symposium on testing in econometrics.

#### 4 CHAPTER 1 ♦ Introduction

- The data may be badly measured or may correspond only vaguely to the variables in the model. “The interest rate” is one example.
- Some of the variables may be inherently unmeasurable. “Expectations” are a case in point.
- The theory may make only a rough guess as to the correct functional form, if it makes any at all, and we may be forced to choose from an embarrassingly long menu of possibilities.
- The assumed stochastic properties of the random terms in the model may be demonstrably violated, which may call into question the methods of estimation and inference procedures we have used.
- Some relevant variables may be missing from the model.

The ensuing steps of the analysis consist of coping with these problems and attempting to cull whatever information is likely to be present in such obviously imperfect data. The methodology is that of mathematical statistics and economic theory. The product is an econometric model.

### 1.3 DATA AND METHODOLOGY

The connection between underlying behavioral models and the modern practice of econometrics is increasingly strong. Practitioners rely heavily on the theoretical tools of microeconomics including utility maximization, profit maximization, and market equilibrium. Macroeconomic model builders rely on the interactions between economic agents and policy makers. The analyses are directed at subtle, difficult questions that often require intricate, complicated formulations. A few applications:

- What are the likely effects on labor supply behavior of proposed negative income taxes? [Ashenfelter and Heckman (1974).]
- 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? (Still to be analyzed.)
- Does attending an elite college bring an expected payoff in lifetime expected income sufficient to justify the higher tuition? [Krueger and Dale (2001) and Krueger (2002).]
- Does a voluntary training program produce tangible benefits? Can these benefits be accurately measured? [Angrist (2001).]

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

The field of econometrics is large and rapidly growing. In one dimension, we can distinguish between theoretical and applied econometrics. Theorists develop new techniques and analyze the consequences of applying particular methods when the assumptions that justify them are not met. Applied econometricians are the users of these techniques and the analysts of data (real world and simulated). Of course, the distinction is far from clean; practitioners routinely develop new analytical tools for the purposes of

the study that they are involved in. This book contains a heavy dose of 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.”

Another loose distinction can be made between microeconomics and macroeconomics. The former is characterized largely by its analysis of cross section and panel data and by its focus on individual consumers, firms, and micro-level decision makers. Macroeconomics is generally involved in the analysis of time series data, usually of broad aggregates such as price levels, the money supply, exchange rates, output, and so on. Once again, the boundaries are not sharp. The very large field of financial econometrics is concerned with long-time series data and occasionally vast panel data sets, but with a very focused orientation toward models of individual behavior. The analysis of market returns and exchange rate behavior is neither macro- nor microeconomic in nature, or perhaps it is some of both. Another application that we will examine in this text concerns spending patterns of municipalities, which, again, rests somewhere between the two fields.

Applied econometric methods will be used for estimation of important quantities, analysis of economic outcomes, markets or individual behavior, testing theories, and for forecasting. The last of these is an art and science in itself, and (fortunately) the subject of a vast library of sources. Though 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. The first 18 chapters of the book are largely devoted to results that form the platform of both areas. Chapters 19 and 20 focus on time series modeling while Chapters 21 and 22 are devoted to methods more suited to cross sections and panels, and used more frequently in microeconomics. Save for some brief applications, we will not be spending much time on financial econometrics. For those with an interest in this field, I would recommend the celebrated work by Campbell, Lo, and Mackinlay (1997). 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.

The techniques used in econometrics have been employed in a widening variety of fields, including political methodology, sociology [see, e.g., Long (1997)], health economics, medical research (how do we handle attrition from medical treatment studies?) environmental economics, transportation engineering, and numerous others. Practitioners in these fields and many more are all heavy users of the techniques described in this text.

## 1.4 PLAN OF THE BOOK

The remainder of this book is organized into five parts:

1. Chapters 2 through 9 present the classical linear and nonlinear regression models. We will discuss specification, estimation, and statistical inference.
2. Chapters 10 through 15 describe the generalized regression model, panel data

**6 CHAPTER 1 ♦ Introduction**

- applications, and systems of equations.
3. Chapters 16 through 18 present general results on different methods of estimation including maximum likelihood, GMM, and simulation methods. Various estimation frameworks, including non- and semiparametric and Bayesian estimation are presented in Chapters 16 and 18.
  4. Chapters 19 through 22 present topics in applied econometrics. Chapters 19 and 20 are devoted to topics in time series modeling while Chapters 21 and 22 are about microeconometrics, discrete choice modeling, and limited dependent variables.
  5. Appendices A through D present background material on tools used in econometrics including matrix algebra, probability and distribution theory, estimation, and asymptotic distribution theory. Appendix E presents results on computation. Appendices A through D are chapter-length surveys of the tools used in econometrics. Since it is assumed that the reader has some previous training in each of these topics, these summaries are included primarily for those who desire a refresher or a convenient reference. We do not anticipate that these appendices can substitute for a course in any of these subjects. The intent of these chapters is to provide a reasonably concise summary of the results, nearly all of which are explicitly used elsewhere in the book.

The data sets used in the numerical examples are described in Appendix F. The actual data sets and other supplementary materials can be downloaded from the website for the text,

[www.prenhall.com/greene](http://www.prenhall.com/greene)