Thad W. Mirer

Economic Statistics and Econometrics

ECONOMIC STATISTICS AND ECONOMETRICS

Thad W. Mirer

State University of New York at Albany

MACMILLAN PUBLISHING CO., INC.
New York

COLLIER MACMILLAN PUBLISHERS
London

Printed in the United States of America

Macmillan Publishing Co., Inc.
866 Third Avenue, New York, New York 10022

Collier Macmillan Canada, Inc.

Library of Congress Cataloging in Publication Data

Mirer, Thad W.
Economic statistics and econometrics.

Bibliography: p.
Includes index.
1. Economics–Statistical methods. 2. Econometrics.
I. Title.
HB137.M57 1983 300'.028 82-17113
ISBN 0-02-381810-7

Printing: 1 2 3 4 5 6 7 8 Year: 3 4 5 6 7 8 9 0 1
ISBN 0-02-381810-7

Preface

This book is designed for courses in economic statistics and introductory econometrics that aim to mix the development of technique with its application to real economic analysis. No background in statistics or calculus is required, and there is no matrix algebra. Nonetheless, the subject matter is technical, and an attempt has been made to work carefully with difficult material rather than to oversimplify it.

There are six parts:

1. Data and Description
2. Specification and Estimation of Regression Models
3. Probability Distributions
4. Inference in Regression
5. Topics in Econometrics
6. Topics in Statistics

The first four parts serve as the basis for a course in ***economic statistics***. Most texts used for this subject give detailed attention to topics in basic statistics and give short shrift to the use of regression in economics. This leaves students with very little to carry on to the rest of their studies. This book is designed for a one-term course with the opposite priorities: only those statistical topics that constitute a foun-

dation for basic econometrics are covered, and regression models are treated in detail.

The book is written to give the instructor considerable latitude in tailoring a course. The more difficult material in these parts, which includes some derivations and the use of logarithms, can be covered selectively or not at all. A fast-paced course can add econometric topics from Part Five, especially when students' backgrounds permit some of the earlier chapters to be covered briefly. A more standard course can add statistical topics from Part Six. The order of Parts Two and Three can be switched, yielding the usual separation of statistics and econometrics.

Parts Two, Four, and Five serve as the basis for a course in ***introductory econometrics***, with Parts One and Three being available for a review of elementary statistics. In recent years several econometrics texts aimed at the senior and beginning graduate levels have appeared. In comparison with them, this book is stronger on the fundamentals of regression and its application but has less theoretical detail and coverage of advanced topics. Hence, this text will be especially useful for courses in which the emphasis is on application or in which students do not have a strong mathematical background.

The outline of the book reveals its special features. In Part One, a fairly unique chapter on the nature of economic data presents two data sets—one cross-section and the other time-series—that serve as the basis for most of the quantitative examples in the book. This is followed by chapters on descriptive statistics and frequency distributions. Part Two consists of three chapters on the specification, estimation, and interpretation of regression models. This is carried out prior to any discussion of statistical inference so that the tasks of specification and interpretation can be considered in detail. A variety of specification forms that can be used in simple and multiple regression are presented, with applications. Part Three begins with an optional chapter on probability theory and then presents the theory of random variables and probability distributions. The normal and t distributions are explained with an eye toward their later use, and an appendix on chi-square and F distributions provides background for special topics appearing later in the book. Part Four develops sampling theory and the methods of statistical inference wholely in the context of regression. This novel presentation focuses attention where it is most critical: on the regression coefficients. Hypothesis testing for single coefficients is developed in a series of special cases using t tests, and confidence intervals are developed for estimation and prediction. Part Five covers a series of topics in econometrics: autocorrelation, heteroscedasticity, F tests, dummy dependent variables, distributed lags, time-series analysis, and simultaneous-equation models. These four chapters are written at a more advanced level than the others.

Part Six covers some standard topics in statistics: sampling and inference for the mean and variance of a random variable, chi-square tests, and a comparison of analysis of variance with regression. The back matter includes a set of statistical tables, the answers to selected problems, and a bibliography.

Computers play an important role in the application of statistical methods, but this book contains no explicit discussion of their use. This makes sense because of the great diversity of systems and programs, their rapid change, and the acceleration of students' computer literacy. Instructors who include computer use as part of the course will find the data sets in Chapter 2 useful: students can replicate and modify the examples presented throughout the text, and new relations can be explored.

Computers played an important role in the production of this book, beyond the obvious statistical calculations. All major drawings of normal, t, chi-square, and F distributions were computer generated (including several three-dimensional figures), with axes and shading traced by hand. The basic statistical reference tables were also computer generated, permitting the construction of a table of P-values for the t distribution.

Several elements of the book's design aim to facilitate learning and classroom discussion. In the first four parts difficult material involving logarithms and derivations is contained in asterisk-marked (*) sections and appendixes; these may be skipped. Problems appear at the end of each chapter, separated by section, and a dagger (†) marks those problems for which answers appear at the back of the book. All display equations are numbered and running heads give the chapter and section numbers, so that material can be located quickly in reading and discussion. The technical notation is kept as consistent and simple as possible.

Finally, a solutions manual for all the problems in the book is available for instructors using the text.

I am grateful to many people for their assistance and contributions. At Macmillan, Chip Price was enthusiastic, encouraging, and helpful from our first contact, and Leo Malek steered the project through production while teaching me about bookmaking. Several reviewers made very useful suggestions, and I am sure they will recognize their impact. These reviewers are as follows: Giorgio Canarella, California State University Los Angeles; Michael Ellis, North Texas State

University; Michael Intriligator, University of California Los Angeles; Nicholas Kiefer, Cornell University; and Michael Panik, University of Hartford. My colleague Terrence Kinal read through the entire manuscript in its next-to-last form, suggesting changes and correcting mistakes. Stephen Rogowski inspired and fostered the typesetting. The illustrations were drawn by Julie O'Brien and Joyce Van Amburgh, under the direction and coordination of Stephen Shapiro. At the very last stage, Yannan Chen provided keen proofreading. Throughout, my wife and children bore the side effects of my work. I used to think it was merely tradition or writers' cant to thank one's family, but now I know better.

Thad W. Mirer

Contents

1 Introduction

When the Nobel Memorial Prize in Economic Science was first awarded, in 1969, it was given to Ragnar Frisch of Norway and Jan Tinbergen of The Netherlands for their pioneering work in econometrics. At the time, few people had heard of the subject, and even fewer knew much about it. Today econometrics is widely recognized as the primary tool of empirical economic analysis.

Put simply, ***econometrics*** involves the development and use of special statistical methods within a framework that is consistent with the ways of economic inquiry. It is an extension of the field of ***statistics***, which deals with techniques for collecting and analyzing data that arise in many different contexts. ***Economic statistics*** involves the application of these general techniques to economic questions.

1.1 The Nature of the Subjects

A time-honored example will illustrate the nature of the subjects and preview the topics covered in this book. In his *General Theory*, John Maynard Keynes developed the concept of an aggregate consumption function as a stable relation between consumer expenditures

and aggregate income. Although the consumption function was only one part of his macroeconomic theory, its elaboration and testing were crucial in the validation and use of Keynes' other ideas. This was one of the first problems on which the then-young field of econometrics cut its teeth.

Econometric analysis starts from a statement of economic theory, whether it be derived from some sophisticated mathematical optimization technique or from some plain reasoning, and develops a ***structural equation*** that specifies how the value of one variable is determined by the values of other variables. In the case of the Keynesian consumption function, the simplest specification is

$$C = \beta_0 + \beta_1 Y + u \tag{1.1}$$

where C stands for consumption and Y stands for income. The parameters β_0 and β_1 are specified to be unknown constants and are referred to as the ***structural coefficients*** in the equation (β is lowercase "beta," the Greek "b"). The ***disturbance*** u may be thought of as a variable reflecting all the factors in addition to income that help determine consumption. These factors may include variables that are relatively unimportant, and therefore not specifically mentioned as determinants of consumption, as well as measurement error and pure chance.

The structural equation serves as a ***model*** of the economic process determining the level of consumption in the economy. As a model, it necessarily abstracts from reality by simplifying the complexity of the true economic process under consideration. It seeks to get down to the essentials. However, many of the techniques of statistics and econometrics are based on the premise that the specified model is an accurate representation of the way the world works. If the model is wrong in its essentials, then all the quantitative and qualitative conclusions that are drawn from the analysis of data using these techniques may be far off the mark. The challenge of econometrics is to blend knowledge of economic theory and behavior with knowledge of statistical techniques in order to produce well-specified models.

Not all equations in empirical economics are structural equations describing economic behavior. For example, the familiar GNP accounting identity

$$Y = C + I + G \tag{1.2}$$

has no unknown parameters and involves no disturbance term. The equation perfectly describes a relation among variables that holds true because of the way the variables are defined and measured. Here, there is no econometric problem.

Returning to the specification (1.1), the main task is to estimate the values of β_0 and β_1 using available data. Applying a technique known as ***regression***, in Chapter 6 we estimate β_0 and β_1 to be 0.568 and 0.907, respectively, based on data presented in Chapter 2. We write the estimated model as

$$\hat{C} = 0.568 + 0.907Y \tag{1.3}$$

where $\hat{C}$ stands for the value of C that is predicted to occur in conjunction with any given value for Y. The result is illustrated schematically in Figure 1.1, where the estimated model is graphed through a scatter plot of data on C and Y. Since the original specification includes a disturbance term, we recognize that the predictions made with this estimated model are always subject to some error. Also, since the numbers 0.568 and 0.907 are only estimates of the true β_0 and β_1, we recognize that there is a further source of error in our predictions.

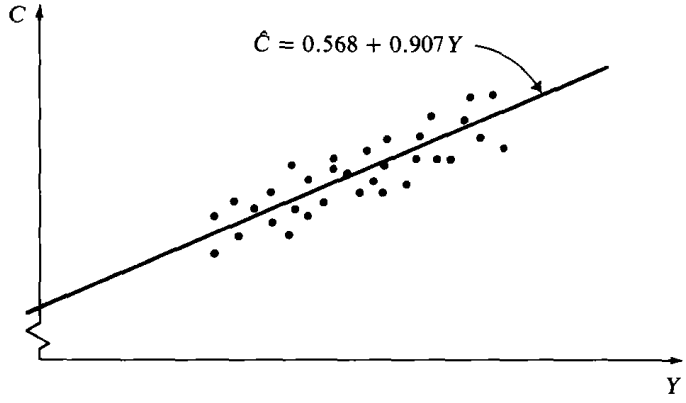

Figure 1.1 Each of the plotted points represents the paired values of consumption (C) and income (Y) in a particular year. The graphed line represents the estimated model of the aggregate consumption function, given as Equation (1.3). (The plotted points are merely illustrative; they are not the actual data points used in the estimation.)

In addition to, or instead of, making predictions we might have special interest in the values of the parameters themselves. For example, in modern technical terms the essence of Keynes' consumption theory is that the marginal propensity to consume is less than 1; this implies that $\beta_1 < 1$. Treating this as a hypothesis to be tested, we note that since our estimated value for β_1 is 0.907, it appears that the hypothesis is confirmed. However, remembering that 0.907 is only an

estimate of β_1, and therefore subject to estimation error, should we really consider the hypothesis to be confirmed? To answer this question, which is related to questions regarding the errors associated with our estimates and predictions, we must gain a firm understanding of the statistical foundations on which econometrics is built.

The field of statistics is divided into two parts, descriptive statistics and statistical inference. ***Descriptive statistics*** is concerned with summarizing the information in data on one or more variables, and it provides the methods for estimating the values of various parameters, including the coefficients of an econometric model. ***Statistical inference*** is concerned with the relation between these estimates and the true values of the parameters, and it provides the basis for testing hypotheses and for assessing the errors that are always present in estimation.

As noted earlier, statistical analysis of economic questions need not be econometric in nature. For example, to estimate the extent of poverty in the United States, the U.S. Bureau of the Census takes an annual survey of more than 50,000 households, asking questions about income and related items. Taking a representative survey involves sampling, and using the results to estimate the proportion of households that are poor involves inference. These techniques are statistical, but not econometric.

1.2 The Plan of This Book

Most readers of this book are students in courses in economic statistics or econometrics, and it has been designed for them. (Naturally, others are welcome, too.) These subjects get taught in a wide variety of ways at different levels, and the book aims to be as flexible as possible. Parts One through Four, with possible deletions, provide the basis for a course in economic statistics. Parts Two, Four, and Five provide the basis for an introductory course in econometrics.

The field of statistics is quite broad and it has a wide variety of applications. This book covers only those fundamental statistical ideas that make up the foundations of econometrics. This coverage is satisfactory at the introductory level, but students planning advanced work in econometrics are well advised to learn statistics on its home ground (i.e., in departments of mathematics and statistics).

Econometrics is a broad field also. This book gives a thorough treatment of the fundamental principles of specification and estimation of single-equation regression models and of the corresponding procedures of statistical inference. The standard "advanced topics" in econometrics are treated in a straightforward way, aiming for appreciation and understanding rather than rigorous detail.