Introduction to ECONOMETRICS

Henri Theil

Introduction to ECONOMETRICS

Henri Theil

University of Chicago

Prentice-Hall, Inc., Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging in Publication Data

THEIL, HENRI.
Introduction to econometrics.

Includes bibliographical references and index.
1. Econometrics. I. Title.
HB139.T47 330'.01 77-14972
ISBN 0-13-481028-7

Printed in the United States of America

10 9 8 7 6 5 4 3 2

PRENTICE-HALL INTERNATIONAL, INC., *London*
PRENTICE-HALL OF AUSTRALIA PTY. LIMITED, *Sydney*
PRENTICE-HALL OF CANADA, LTD., *Toronto*
PRENTICE-HALL OF INDIA PRIVATE LIMITED, *New Delhi*
PRENTICE-HALL OF JAPAN, INC., *Tokyo*
PRENTICE-HALL OF SOUTHEAST ASIA PTE. LTD., *Singapore*
WHITEHALL BOOKS LIMITED, *Wellington, New Zealand*

Preface

The main objective of this book is to enable the reader to understand applied econometric studies. The prerequisites for reading the book consist of elementary calculus through partial differentiation, apart from high school algebra and one or two courses in economics; further details are provided on page xvii. A second objective is to provide more advanced material in the sections marked by an asterisk. These sections can either be used for supplementary reading or for the second course in a two-course sequence.

Acknowledgments

I have used several preliminary versions of this book in the introductory econometrics course for MBA and college students in Chicago. Their reactions were a major determinant of the present form of the book. I am also indebted to Professors Walter D. Fisher (Northwestern University), Robert McNown (University of Colorado), and Houston Stokes (University of Illinois), who made detailed comments on an earlier draft; to Kenneth Laitinen, who provided comments on successive drafts while assisting in the teaching of the course; to Professors Jacques Drèze (University of Louvain) and Teun Kloek (Erasmus University, Rotterdam) and my colleagues William Wecker and Victor Zarnowitz, who commented on parts of the book; and to Prentice-Hall's Frank Enenbach, who is *almost* on our faculty and persuaded me to give the book to his firm. I want to express my great appreciation for the hard work and cheerfulness of my secretary, Mrs. Janice Ricks, and her predecessor, Mrs. Kathleen Masulis, who typed and retyped and checked and rechecked galleys and page proofs.

The Institute of Mathematical Statistics kindly permitted me to include the table of the Von Neumann ratio; so did the Biometrika Trustees for the t, χ^2, F, and Durbin-Watson tables. These Trustees and Longman Group Ltd. informed me that some columns of the t table are from *Statistical Tables for Biological, Agricultural and Medical Research* by Fisher and Yates (sixth edition, 1974). I am grateful to the Literary Executor of the late Sir Ronald A. Fisher, F.R.S., to Dr. Frank Yates, F.R.S., and to Longman Group Ltd., London, for permission to reprint these columns.

HENRI THEIL
Chicago, Illinois

The Design and Use of This Book

Each chapter consists of one or more opening paragraphs and several sections (e.g., Section 3.2 is the second section of Chapter 3). Most sections contain an opening paragraph and several (unnumbered) subsections. Formulas are indicated by two numbers, the first for the chapter and the second for the order of occurrence; for example, eq. (5.3) is the third equation in Chapter 5. Tables and figures are numbered consecutively throughout the book. To facilitate finding a table or figure, its page number is usually mentioned.

The Appendix consists of eight sections, labeled A to H. Formulas in the Appendix are indicated by the letter of the section followed by the number of the order of occurrence.

Prerequisites

Elementary calculus is a prerequisite for reading this book, but analytical integration is not used. The reader can work his way through the material on continuous distributions in Sections 4.1 and 4.2 by simply remembering that the probabilities of discrete distributions (Chapter 3) are replaced by the density function and summation by integration.

Apart from the elementary calculus prerequisite, the book presumes no further mathematical or statistical preparation. Statistics is developed from scratch. The use of the summation operator $\sum$ is explained in Section 2.2. On the other hand, if the reader has a sufficient command of matrix algebra, he will appreciate the concise style of the Appendix. No attempt is made to explain matrix algebra in this book, but an instructor who uses the book for a two-course sequence will not find it difficult to provide the necessary material for Sections A and B of the Appendix in his second course.

The Four Parts of the Book

Since the book is introductory, it should emphasize the standard linear model. Accordingly, the main objective of Part One is this model for one explanatory variable, that of Part Two is the extension for more variables and the associated statistical inference techniques, whereas the discussion of such topics as generalized and two-stage least squares is postponed until Chapters 19 to 21 at the end of Part Three. In a book of this kind it is impossible to describe even a medium-sized simultaneous-equation model adequately; the familiar Klein Model I must do the work. To provide a compensation, I decided to discuss predictions and prediction errors in some detail, as well as certain elements of control theory, in Chapters 22 and 23, which constitute Part Four.

Given the presumption that the reader knows little or no mathematical statistics, it would be a disadvantage to have all statistics in one or several chapters at the beginning. Few persons would be able to digest all this material. Therefore, statistics is introduced successively in the first three parts of the book. Chapters 3 and 4 in Part One describe discrete and continuous distributions and the first elements of point estimation theory; this statistical basis is sufficient through Chapter 9. The second statistical installment is Chapters 10 to 12 in Part Two, dealing with multivariate normality, χ^2, t, confidence intervals, and hypothesis testing. The third is Chapters 15 and 16 at the beginning of Part Three on large-sample distribution theory, Bayesian inference, and related matters. This division of the statistical material into successive stages has obvious additional advantages when the reader is a student who takes an introductory mathematical statistics course concurrently.

An important question is at which stage the reader should be advised to be suspicious. Should he be told each time immediately after a new approach is introduced how it can go wrong because the data may violate the underlying assumptions? Or should these warnings be postponed until the stage at which he has developed some feeling for the relationship among the various approaches? I prefer the latter alternative, since the former runs the risk of generating confusion rather than understanding. For example, it would be very troublesome to explain the consequences of an erroneous specification of the relations estimated when properties of least squares are discussed in Chapters 5 and 8. Accordingly, doubts are spread in Chapter 17, which deals with specification and observational errors, and in Chapter 18 (on the consequences of aggregation) and in Section 23.2 (on economy-wide econometric models).

Using the Book in the Classroom

The econometrician analyzes statistical data. Thus, when an instructor teaches an introductory econometrics course, he should expect that his students will want to see some "real" data at an early stage. But if these students know little or no statistics, there is the risk that they will spend weeks wading through algebraic symbols before they see anything "real." How can this problem be resolved?

I find it useful to divide my class sessions (about 75 minutes each) into two or three parts, so that several topics can be discussed in each session and one of these can be data analysis. An additional advantage of this division-into-parts is that it reduces the accumulation of mathematical results, so that it avoids a situation in which students are asked to apply results which they have not fully digested.

For example, in the first session the instructor may want to start with the first few pages of Section 1.1, which describe (among others) a consumption-income relation, and then move to Section 2.1, where such a relation is considered for fictitious data. He assigns Section 2.2 for the second session and moves on to Section 2.3 until the discussion of the least-squares principle, eq. (2.25). He then proceeds to the opening paragraph of Chapter 7, which contains real data (Table 12), and discusses the questions raised in that paragraph. The last topic discussed in the first session is Section 3.1 on univariate discrete distributions. In the next two sessions the instructor completes Chapters 2 and 3. As soon as Chapter 2 is completed, he goes through Section 7.1, where least squares is applied to the data of Table 12. This means that at an early stage of the course students have seen how data are used to obtain an income elasticity and a price elasticity. Further details on this course are provided in the Teacher's Manual.

The application of econometric methods to actual economic data is a prominent feature of this book. There are two ways in which applications can be presented: Either numerous examples are discussed briefly, or few examples are discussed in more detail. I prefer the latter alternative, because numerous brief discussions easily lead to a superficial attitude, which is unfortunately too common in applied econometric studies. Accordingly, the above-mentioned data of Table 12 figure prominently in the confidence intervals and statistical tests of Chapters 11 and 12, and Klein's Model I plays an important role in several sections of Chapters 21 and 23. Obviously, discussing few examples has a disadvantage also, but this can be eliminated to a large degree by means of a display of scatter diagrams referring to different areas of application. Several such diagrams are mentioned in the opening paragraph of the References at the end of this book.

Chapters 14 and 22 should be viewed as contrasts. Chapter 22 is mainly empirical and serves to provide the student with an idea of the degree of accuracy which he can expect from economic predictions and of the decline of this accuracy when the prediction refers to a more distant future. Chapter 14 provides a link between economic theory and econometric analysis. The economic theory used in this chapter is that of utility maximization subject to a budget constraint, which yields Slutsky symmetry. Usually, this symmetry is viewed as no more than an interesting academic result, but it has practical implications. Ignoring income effects (which are usually small), the instructor can describe Slutsky symmetry for his students as follows: A one-dollar increase in the price of your product has an effect on the demand for your competitor's product which is equal to the effect on the demand for your product of a one-dollar increase in the price of your competitor's product. When Slutsky symmetry is described in this way, students will be better motivated to derive it theoretically, test it statistically, and impose it on the estimates.

Starred Sections

From Chapter 10 onward, certain sections (sometimes all sections) of most chapters are marked by an asterisk. Such sections can be omitted in a first reading; later sections without asterisks are independent of earlier sections with asterisks except for an occasional sentence in parentheses or brackets. The instructor can delete the starred sections in an elementary course, although the allocation of stars is obviously arbitrary to some degree.

Deleting starred sections in an introductory course is particularly attractive when this course is part of a two-course sequence: The instructor can then ask students to review the earlier sections, discussed in the first course, before he proceeds to the later sections of the same chapter in the second course. However, even if the sections without asterisks are discussed in one single course, without a second-course follow-up, the students will have a fairly detailed knowledge of the standard linear model, including confidence intervals and t tests, some knowledge of simultaneous-equation models and of specification, heteroscedasticity, and autocorrelation problems, as well as an appreciation of the analysis of numerical economic data in conjunction with the corresponding economic theory.

Contents

part one
THE SIMPLE REGRESSION MODEL

chapter 8

Multiple Regression 110

chapter 9

Multiple Correlation 128

chapter 10

Multivariate Normality and Maximum Likelihood 140

chapter 11

Confidence and Prediction Intervals 153

part three
EXTENSIONS OF THE REGRESSION MODEL

chapter 18

Aggregation 288

chapter 19

Generalized Least Squares and Related Procedures 299

chapter 20

Introduction to Simultaneous-Equation Problems: Static Equation Systems 319