ECONOMETRICS

Maddala

CONOMICS HANDBOOK SERIES

ECONOMETRICS

G. S. Maddala
University of Florida

McGraw-Hill Publishing Company
New York St. Louis San Francisco Auckland Bogotá
Caracas Hamburg Lisbon London Madrid Mexico Milan
Montreal New Delhi Oklahoma City Paris San Juan
São Paulo Singapore Sydney Tokyo Toronto

ECONOMETRICS

14 15 BRBBRB 90

This book was set in Times Roman by Computype, Inc. The editors were J. S. Dietrich and Michael Gardner; the production supervisor was Charles Hess. The drawings were done by Vantage Art, Inc.

Library of Congress Cataloging in Publication Data

Maddala, G S date
Econometrics.

(Economics handbook series)
Includes index.
1. Econometrics. I. Title.
HB139.M35 330'.01'82 76-26042
ISBN 0-07-039412-1

PREFACE

Many students at universities and applied econometricians in government and industry have expressed the opinion that most of the books in econometrics now available concentrate on theory and a limited number of topics. Thus they have to consult different books for different purposes. They felt a need for a comprehensive book with an empirical bias. The present book has been written with these people in mind. It was written entirely during 1974–75 and has not had the benefit of being tried in an actual teaching environment; otherwise, the text might have taken a different form.

Though I have learnt a lot from the work of all the respected econometricians in the profession (and thanking them all would make a long preface), the practical orientation of the book is due to the influence of Art Goldberger, Zvi Griliches, and Marc Nerlove, the first of whom I had contact with mainly through papers. Also my association with Jaques Dréze (since reading his pioneering paper of 1962) and Arnold Zellner has convinced me that every student in econometrics needs an exposure to the Bayesian approach. None of these people, however, have read any part of the book and so cannot be blamed for any shortcomings.

I would like to thank my students David Grether, Kajal Lahiri, and Forrest Nelson who went through the book and suggested some corrections. I would also like to thank Ken Gaver and Walter Oi at the University of Rochester for going through the manuscript. None of them are responsible for any remaining errors.

The book started out with undergraduates in mind and ended up at the graduate level. However, there is a steady progression in the level and the beginning portions can be used by undergraduates, whereas the book as a whole can be used by graduate students as well as applied econometricians in government and industry.

Part two of the book "Introduction to Probability and Statistical Inference" is perhaps very short, but elaborating on it would have made the book much too long. I thought of deleting it completely but finally felt that it is useful to have the review, even if it is cursory.

One objective throughout has been to minimize the algebraic detail. Some purists in theory might find portions of the book not sufficiently rigorous, but I had to do this purposefully. Proofs of theorems have been relegated to appendices. Matrix notation occurs only in the last chapters and the Appendixes. Also, at several places in the book the maximum-likelihood method is described, but the derivation of (asymptotic) standard errors is not presented because explicit presentation of the matrix of second derivatives in each and every case would merely clutter up the book with unnecessary notation. Readers can refer to Sec. 9.10 and work out the necessary first and second derivatives in each case.

The last chapter on Bayesian inference in econometrics may appear unsatisfactory to some "Bayesians." My purpose in including it was to emphasize the similarities between the classical and Bayesian approaches. Further, it gives a review of the area for those who do not want to spend time (initially) reading specialized books in the field. Likewise, some readers might find the discussion on limited and qualitative variables in Chap. 9 very brief. Again the purpose has been to expose readers to the main problems in this area. Since the field is vast, doing full justice to all the problems would have made the book unusually long.

Appendixes A, B, and C contain almost all the material in matrix notation that is often covered in graduate courses in econometrics at many universities. The exercises in Appendix D are drawn from some econometrics examinations at Chicago, Florida, Rochester, Stanford, and Yale.

I would like to thank Marjorie Adams, Martha Colburn, Carroll Cornwall, Susan Groth, and Janet Wood at the University of Rochester for their careful typing of the manuscript.

Finally, I would like to thank my wife, Kameswari, my daughter, Tara, and my son, Vivek, for their encouragement toward the completion of this book.

G. S. Maddala

CONTENTS

PART ONE

INTRODUCTION

CHAPTER

ONE

DATA, VARIABLES, AND MODELS

Econometrics consists of an application of statistical methods to economic data. However, some special problems associated with economic data and economic relationships necessitate a separate discussion of these methods. In this book we will be illustrating these problems with reference to economic models and economic data. Before we proceed, we need to discuss briefly the special problems peculiar to economics. These can be classified under the headings:

1. Data
2. Relationships
3. Variables
4. Functional forms

1-1 DATA

The data we observe in economics are of two types: cross-section data and time-series data. In cross-section data we have observations on individual units at a point of time, e.g., data on consumer income and expenditures on food for a set of families, data on teacher salaries and characteristics, data on labor-force participation, or wages and characteristics of workers. These data are usually collected by some sample surveys. Hence before using these data, one should examine the type of survey conducted. Another type of cross-section data that is often used is a cross section of states or of regions. There have been many studies on demand functions, production functions, and cost functions in which the individual observations are aggregates over states. In fact there are several cross-country, cross-section studies in which the individual observations are the aggregates for different countries. This is perhaps stretching the definition of a cross section too far.

In time-series data we have observations over a period of time, e.g., quarterly data on GNP or monthly data on industrial production or employment. Economics involves a lot of time-series data relative to other fields. However, these time-series are often very short, and most of the series move up and down together. Hence there are severe problems in inferring cause and effect. In econometrics most theoretical developments have been in devising methods of handling time-series data, and often what one finds is an application of these sophisticated techniques to scanty data. Sometimes we find attempts to increase the number of observations, e.g., by using monthly series rather than quarterly series. But these create special problems, and moreover the monthly series is often obtained by an interpolation of the quarterly series. Another device used to increase the number of observations is considering a combination of cross-section and time-series data, e.g., if we have data on sales, profits, and investment for a number of companies over a number of years or if we have data on gasoline consumption, number of cars, population, income, etc., for a number of states over a number of years.

1-2 RELATIONSHIPS

The relationships we investigate in economics are of three types:

1. Single-equation
2. Multiple-equation
3. Simultaneous

In single-equation relationships there is a *dependent* or "determined" variable which is determined by one or more *independent* or "determining" variables; e.g., when we say consumption depends on income Y, wealth W, and rate of interest r, we write $C=f(Y,W,r)$. Here C is the dependent variable and Y, W, and r are the independent variables. If we are considering a family with given income, wealth, and market rate of interest, this relationship can be used to determine how the consumption C of the family changes in response to changes in these variables. Similarly, when we say quantity demanded depends on price, we can write $Q=f(P)$. Here Q is the dependent variable and P the independent variable. If we are faced with an individual customer faced with a given price, this equation can be used to determine how the quantity he purchases Q changes with changes in the market price P.

In multiple-equation relationships we have a set of equations. For example, let C_A, C_D, C_{ND}, and C_S denote, respectively, consumer expenditures on automobiles, on other durables, on nondurables, and on services. Each of these could be a function of income and wealth. But the way they depend on these variables could be different. Hence, instead of studying the relationship between total consumer expenditures and income and wealth, we would gain more knowledge by studying the relationship between C_A, C_D, C_{ND}, and C_S and income and wealth. We now have a four-equation system. In such cases sometimes we can treat each of these equations separately as in a single-equation relationship. But sometimes we have to treat them together.

In simultaneous-equation relationships, two or more variables are determined "simultaneously" by a number of determining variables. In the above examples, though Y is "given" for the individual family, for the economy as a whole we cannot treat Y as being "given." We have to treat both C and Y as being determined simultaneously by some policy variables and technological and sociological conditions. Similarly, price can be treated as "given" for the individual customer, but if we are considering the market as a whole, we have to consider both price and quantity as being determined simultaneously by demand and supply conditions and other variables. In this case we have a two-equation system like

$$Q = f(P,X) \qquad \text{demand relationship}$$

$$Q = g(P,Z) \qquad \text{supply relationship}$$

These two equations together determine Q and P, given the determining variables X and Z. For example, X could be income and Z could be weather. Simultaneous-equation relationships are also multiple-equation relationships. But there is an essential difference in the way the variables are interconnected. These differences will be clear when we discuss these problems in subsequent chapters.

1-3 VARIABLES

A common terminology used in econometrics for dependent and independent variables is *endogenous* and *exogenous* variables, respectively. Endogenous variables are those determined within the economic system, and exogenous variables are those given from outside the system. In a broad sense almost all variables are endogenous and the only exogenous variables one can think of are weather, cyclones, etc. However, in any problem this is a matter of approximation. While studying the demand for gasoline by households, we can treat the quantity demanded as endogenous and income and price as exogenous, arguing that the household does not have control over these. Similarly, for some purposes we can treat government expenditures and taxes as exogenous. However, as we lengthen the time period of our observations, these variables will also become endogenous. In general, the greater the level of aggregation—whether it be over time periods or over individual cross-section units—the more exogenous variables will have to be treated as endogenous.

Endogenous variables can further be classified as target and nontarget variables. Target variables are those we like to influence. Nontarget variables are those we do not care about; e.g., employment and price level may be target variables. Similarly, exogenous variables can be classified as instruments and noninstruments. An instrument is an exogenous variable that is specifically manipulated so as to achieve some targets. Government expenditures, taxes, and subsidies are examples of instruments.

1-4 FUNCTIONAL FORMS

Economic theory may tell us that quantity demanded is a function of price, but it may not tell us the functional form of the relationship; e.g., is the relationship of the form:

Linear: $Q = \alpha + \beta P$
Log-linear: $\log Q = \alpha + \beta \log P$
or semilog: $\log Q = \alpha + \beta P$
or: $Q = \alpha + \beta \log P$?

This is something we decide on the basis of the observed data. The observed data on Q and P can be graphed to see what the relation looks like. If the observations are as in Fig. 1-1, a linear relationship is appropriate. If they are as in Fig. 1-2, we need a curvilinear relationship to describe the data. In this case we can plot $\log Q$ against $\log P$, $\log Q$ against P, and $\log P$ against Q and see which of these is approximately linear. However, this is a very simplistic description of what should be done. Often there may be cases where none of these simple functional forms work well. For example, consider the adoption of a new product. Usually, it has been observed that such phenomena are described by a process as in Fig. 1-3. The rate of adoption is slow in the beginning stages; then it picks up, and finally it tapers off. In such cases one can describe

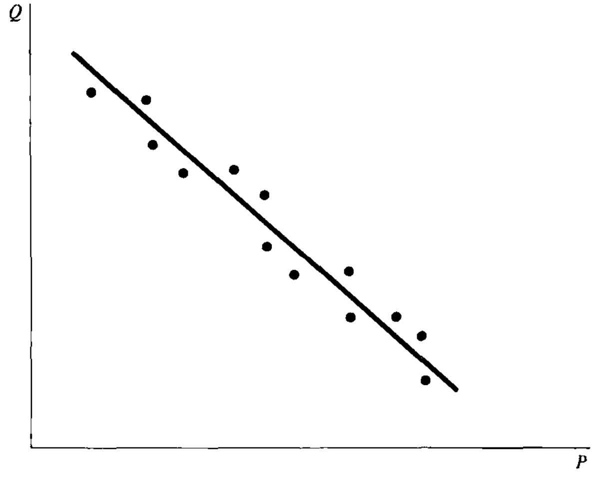

Figure 1-1 Linear relationship.

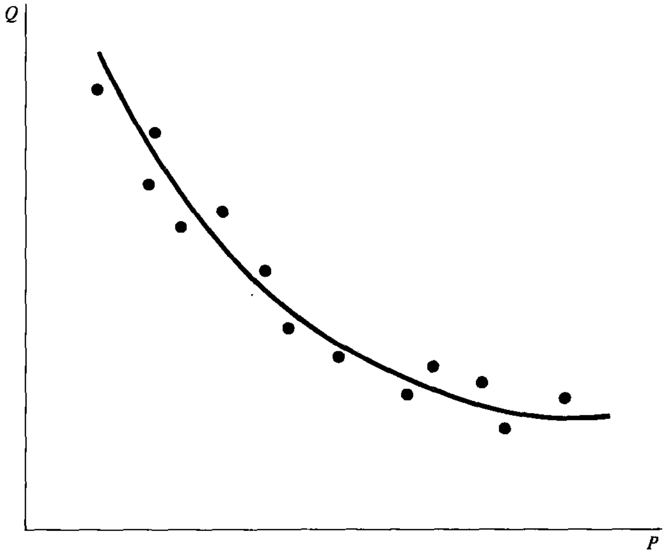

Figure 1-2 Curvilinear relationship.

the data by what is called a *logistic function*. It is given by

$$P = \frac{c}{1 + ae^{-bt}} \qquad (a > 0,\ b > 0,\ c > 0)$$

Here, P is the proportion of households who have adopted the product at time t and c is the ceiling of this proportion. As $t \to \infty$, $P \to c$. The higher the value of b, the higher the rate of approach of P to the ceiling c.

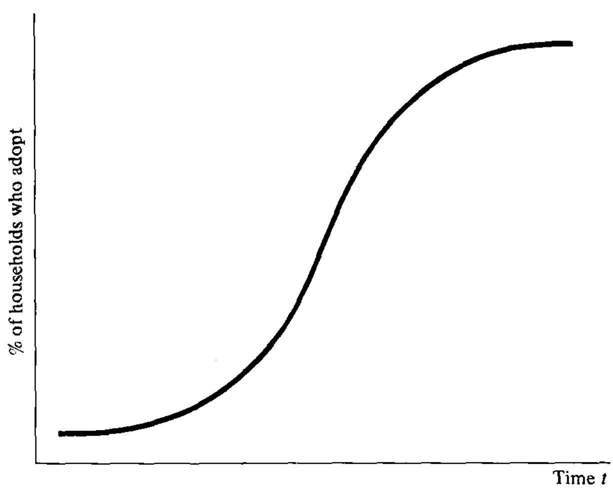

Figure 1-3 Logistic curve.