ECONOMETRIC METHODS

Econometric Methods

J. Johnston
Professor of Econometrics
University of Manchester, England

McGraw-Hill Book Company, Inc.
New York · San Francisco
Toronto · London

ECONOMETRIC METHODS

Library of Congress Catalog Card Number 62-17369

7 8 9 - M P - 9 8 7

32680

Econometric Methods

By the same author

Statistical Cost Analysis
(McGraw-Hill, New York, 1960)

For A., R., and M.

Preface

The purpose of this book is to provide a fairly self-contained development and explanation of econometric methods for students who have already done about one year's work in statistical theory and method. It is divided into two parts. Part 1 contains a full exposition of the linear normal regression model. This serves as an essential basis for the theory of econometrics in Part 2. This latter part expounds the main statistical methods now available for the estimation of econometric models.

Students who have already done a year's work in mathematical statistics will be able to skip much of the first two chapters, which have been inserted as a link with the conventional courses in statistical methods taken by most students in the social sciences. Chapter 1 is a complete treatment of the two-variable linear model, including all the problems of estimation, hypothesis testing, and forecasting which arise in the context of this model. Most students will already be familiar with many of the topics treated in this chapter. An elementary knowledge of probability distributions, expected values, estimation, and hypothesis testing is assumed. Anyone who experiences difficulty with the material in Chap. 1 should refer to a good introductory book on statistics before proceeding any further.[1] As well as providing useful review material, Chap. 1 introduces all the basic inference problems, which will be considered in more complicated contexts throughout the rest of the book.

[1] For example, P. G. Hoel, *Introduction to Mathematical Statistics*, 2d ed., Wiley, New York, 1954. A. M. Mood, *Introduction to the Theory of Statistics*, McGraw-Hill, New York, 1950. D. A. S. Fraser, *Statistics: An Introduction*, Wiley, New York, 1958.

Chapter 2 deals with extensions of the two-variable linear model to embrace nonlinearities and also increases in the number of variables. One does not travel far along this road before the notational and other complexities call for a more powerful technique. Chapter 3 thus provides the essentials of matrix algebra, which is then used as the basic method of exposition throughout the rest of the book. Chapter 4 deals with the general linear model in k variables. This is the basic and final chapter in Part 1, and it contains a development of all the important results for this model. Chapters 1, 2, and 3 provide preparatory material which different readers will have to use more or less intensively before tackling Chap. 4. The material in Chap. 4 is an essential prerequisite for understanding the developments in econometric methods described in Part 2.

After a short introductory chapter, Part 2 contains a treatment of errors in variables in Chap. 6. This is a topic which sometimes receives scant attention in econometric work, but which is often of great practical importance. Chapter 7 gives a unified treatment of problems arising from autocorrelated disturbances, and various other problems which can arise in a single equation context such as multicollinearity, heteroscedasticity, lagged variables, and dummy variables are covered in Chap. 8. The final two chapters deal with simultaneous-equation problems, including identification problems, indirect least squares, two-stage least squares, limited-information methods, full-information, and three-stage least squares.

The emphasis throughout is on the rationale of the various methods. I have attempted to explain as fully as possible the assumptions underlying the various techniques and to give a fairly extensive development of the various results, in the hope that readers of varying backgrounds will be able to work through the material on their own and develop a real appreciation of the advantages and limitations of the various techniques in different practical applications. Numerical examples are given in the text, and theoretical and numerical exercises at the end of most chapters. I am grateful to various authorities for permission to use examples from examination papers of the Royal Statistical Society and the Universities of Cambridge, London, Manchester, and Oxford. I have given no treatment of computational problems since an ever-increasing number of research workers are now using various electronic computers with associated programs.

I am heavily indebted to certain individuals. The project would not have been undertaken at all had it not been for the encouragement and support that I received at a crucial stage from Prof. Guy H. Orcutt at the University of Wisconsin. J. Parry Lewis of the University of Manchester checked through the algebra, and he and R. J. Ball, also of Manchester, made many valuable suggestions. I am very grateful to them both. I am also greatly indebted to Prof. A. S. Goldberger of the University of Wisconsin for sending me his mimeographed lecture notes, which have materially improved my exposition at several points. Thanks also go to Profs. R. G. Lipsey and W. M. Gorman of London and Oxford, respectively, for valuable comments. Miss Pamela Drake checked the numerical examples, and L. T. Simister helped with the survey of empirical studies in Chap. 10. It is also a pleasure to acknowledge the patience and skill with which Mrs. Katherine Norrie and Miss Pauline O'Brien typed several versions of this manuscript. The final burden of proof correction was greatly eased by assistance from David Bugg.

J. Johnston

Contents

PART 1

The Linear Normal Regression Model

1

The Two-variable Linear Model

1-1. Relationships between Variables

The first basic idea to which the student of economics is introduced is that of relationships between economic variables. The quantity demanded of a commodity in a market is regarded as a function of its price, the costs of producing a product are assumed to be a function of the amount produced, consumption expenditure is taken as a function of income, and so forth. These are all examples of two-variable relations, but more realistic formulations require the specification of several variables in each relation. Thus quantity demanded may be regarded as a function of price, disposable income, and prices of related commodities; production costs will depend on rate of production, factor prices, and changes in production rate; and consumption expenditure may be specified as a function of income, liquid assets, and previous consumption levels. Economic theory consists of the study of various groups or sets of relations which are supposed to describe the functioning of a part or the whole of an economic system. The task of econometric work is to estimate these relationships *statistically*, and this empirical testing and measurement of economic relationships is an essential step in the acquisition of economic knowledge.

1-2. The Two-variable Linear Model

The first step in the measurement of economic relationships is the specification of which variables enter the various relations. To simplify matters as much as possible, we shall examine first of all the most elementary case, by making the twin assumptions that we are dealing only with a single relation and that it contains only two variables. Denoting the variables by Y and X, we may postulate

$$Y = f(X) \tag{1-1}$$

This step merely identifies the variable X, which is thought to influence the other variable Y.

The second step is to specify the form of the relation between Y and X. The theory underlying the development of (1-1) may suggest the precise functional form to use, or it may merely suggest certain side conditions on the intercept, slope, and curvature of the function. Such conditions may be satisfied by a variety of functions, and we then look to statistical analysis for some help in choosing between them.

The simplest relationship between two variables is a linear one, namely,

$$Y = \alpha + \beta X \tag{1-2}$$

where α and β are unknown parameters indicating the intercept and slope of the function. Other relationships between two variables include

$$Y = \alpha e^{\beta X} \qquad Y = \alpha X^{\beta} \qquad Y = \alpha + \beta \frac{1}{X}$$

The third relation is linear in the variables Y and $1/X$, and the first and second can be reduced to a linear form in transformed variables by taking logs of both sides to give

$$\log_e Y = \log_e \alpha + \beta X$$

and

$$\log_e Y = \log_e \alpha + \beta \log_e X$$

respectively. The first is linear in log Y and X, and the second is linear in the logs of both variables.

The bulk of conventional economic theory, whether expressed in diagrammatic or algebraic form, postulates exact functional relationships between variables. The most elementary acquaintance with economic data, however, indicates that points do not lie exactly on straight lines or other smooth functions. Thus, for measurement and testing purposes, formulations such as (1-1) and the various functional forms associated with it are inadequate. The extension employed is the introduction of a stochastic term into economic relationships.

Suppose, for example, that we are investigating the relationship between consumption expenditure and disposable income in a cross section of households for some given period in time. Letting Y denote consumption expenditure and X denote disposable income, completed budget data for, say, 10,000 households would provide 10,000 pairs of associated measurements X_i, Y_i ($i = 1, 2, \ldots, 10{,}000$). Let us suppose that we have already divided our households into various groups on the basis of household size and composition and are looking at the relationship between Y and X *within* a given group. We do not expect that all households within the group which have some given income X' will display an identical consumption expenditure Y'. Some will spend more than others, some will spend less, but we do expect a clustering of the expenditure figures around a value which is geared to the income value in question. These ideas may be expressed more formally in a new linear hypothesis

$$Y = \alpha + \beta X + u \tag{1-3}$$

where u denotes a variable which may take on positive or negative values. Thus, if we consider the subgroup of those households with a given income X', the central value of consumption expenditure for them will be $\alpha + \beta X'$, but actual consumption figures for individual households in the subgroup will be indicated by $\alpha + \beta X' + u_1$, $\alpha + \beta X' + u_2$, etc., where $u_1, u_2, \ldots$ indicate the amount by which the expenditures of particular households exceed or fall short of the central value $\alpha + \beta X'$.

There are three possible, though not mutually exclusive, ways of rationalizing the insertion of the u term in (1-3). First, we may say that the consumption expenditure of each and every household could be fully explained if we knew all the factors at work and had all the necessary data. Even among households of the same size and composition, there will be variations in the precise ages of the parents and children, in the number of years since marriage, in whether the husband is a golfer, drinker, poker player, or bird-watcher, in whether the wife is addicted to spring hats, Paris fashions, swimming pools, or foreign sports cars, in whether the household income has been increasing or decreasing, in whether the parents are themselves the children of thrifty, cautious folk or of carefree spendthrifts, and so forth. In explaining human behavior the list of relevant factors may be extended ad infinitum. Many of the factors, however, will not be quantifiable, and even if they