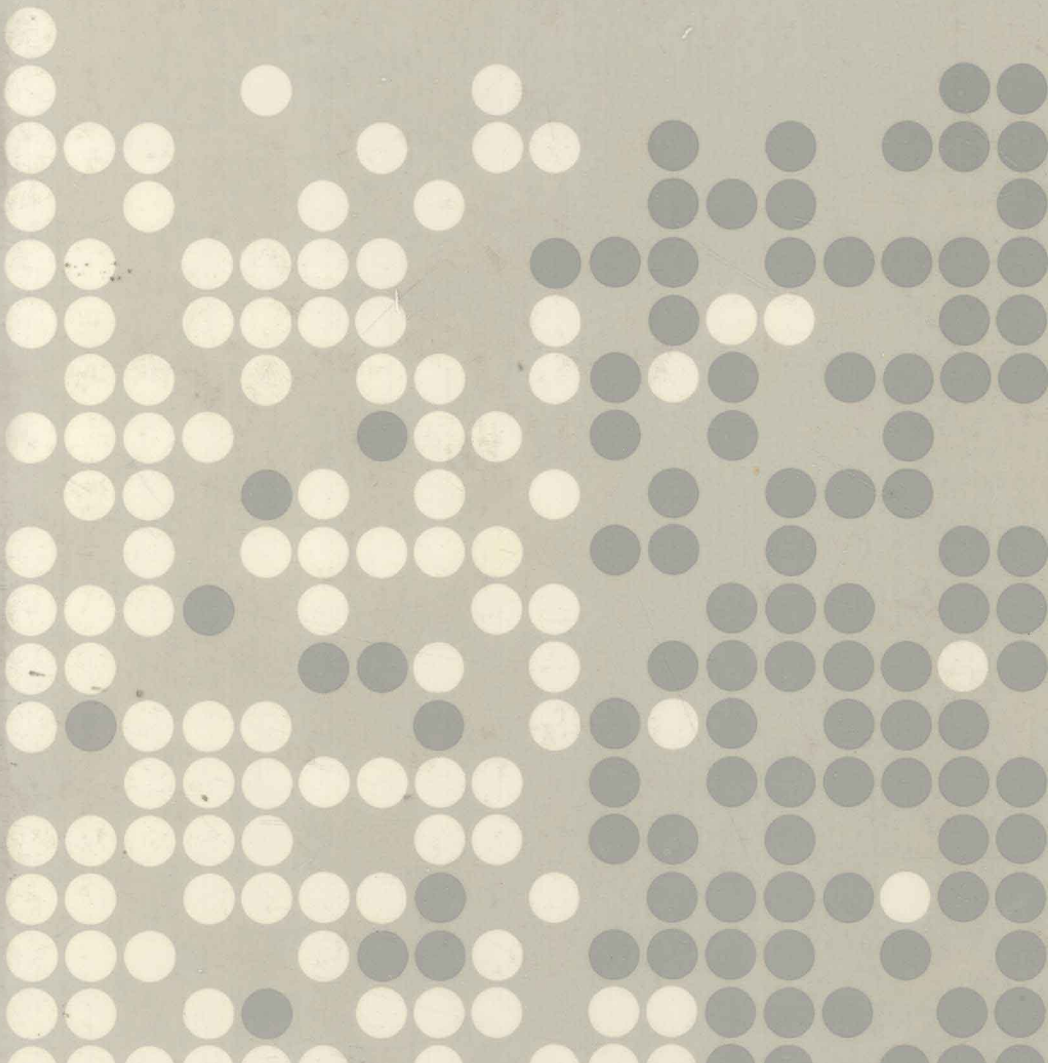
G.S. Maddala

# INTRODUCTION TO ECONOMETRICS



# INTRODUCTION TO ECONOMETRICS

G. S. Maddala
University of Florida

**Macmillan Publishing Company**
**New York**

Collier Macmillan Publishers
London

Printed in the United States of America



Macmillan Publishing Company
866 Third Avenue, New York, New York 10022

Collier Macmillan Canada, Inc.

Library of Congress Cataloging in Publication Data

Maddala, G. S.
Introduction to econometrics.

Includes index.
1. Econometrics. I. Title.
HB139.M353 1988 330'.028 87-32638
ISBN 0-02-374530-4

Printing: 1 2 3 4 5 6 7 8 Year: 8 9 0 1 2 3 4 5 6

# Preface

There have been many important developments in econometrics during the last two decades, but introductory books in the field still deal mostly with what econometrics was in the 1960's. The present book is meant to familiarize students (and researchers) with some of these developments, explaining them with very simple models without cluttering up the exposition with too much algebraic detail. Where proofs involve complicated expressions they are omitted and appropriate references are given. Ten of the more difficult sections have been marked with an asterisk indicating that they are optional. Beginning students can skip them and proceed. The book also contains several examples illustrating the techniques at each stage, and where illustrations are not given, some data sets have been provided so that students can compute the required results themselves.

Since the book covers quite a few topics, only a few examples have been given to illustrate *each* point. Giving too many illustrations for a single point might be boring for some students and would have made the book much too bulky. The study guide contains more illustrative examples, data sets, and questions and answers. The exercises given at the end of each chapter are somewhat challenging. However, the study guide contains answers or guidelines. The study guide also gives a "guided tour" of the material in each chapter as well as some detailed explanations for some points that are touched briefly in the book.

Some of the questions at the end of the chapters have been taken from the examinations at several U.S. and U.K. universities, and from P. C. B. Phillips and M. R. Wickens, *Exercises in Econometrics* (Cambridge, Massachusetts, Ballinger Publishing Co., 1978), Vol. I. (Many of the questions in that book are from examinations in the U.K. universities.) Since questions tend to get repeated with minor changes, I have not bothered to quote the exact source for each question.

There are many distinguishing features of the book, some of which are:

1. A discussion of aims and methodology of econometrics in Chapter 1.
2. A critique of conventional significance levels in Chapter 2.
3. A discussion of direct versus reverse regression and inverse prediction in Chapter 3.
4. A thorough discussion of several practically useful results in multiple regression (Chapter 4), some of which are not to be found even in specialized books on regression.

5. Discussion of several new tests for heteroskedasticity. Choice of linear versus log-linear models, and use of deflators (in Chapter 5).
6. A thorough discussion of the several limitations of the Durbin-Watson test, showing that it is almost useless in practice, and a discussion of Sargan's test, ARCH test, LM test, etc. (Chapter 6).
7. A critical discussion of the use and limitations of condition numbers, variance inflation factors (VIF's), ridge regression, and principal component regression in the analysis of multicollinearity (Chapter 7). Many of these techniques are often used uncritically, because computer programs are readily available.
8. The use of dummy variables to obtain predictions and standard errors of predictions, the relationship between discriminant analysis, and the linear probability model, the logit, probit and tobit models (Chapter 8).
9. Inference in under-identified simultaneous equation models, criteria for normalization and tests for exogeneity and causality (Chapter 9).
10. The discussion of partial adjustment models, error correction models, rational expectations models, and tests for rationality (Chapter 10).
11. Reverse regression, proxy variables (Chapter 11).
12. Different types of residuals and their use in diagnostic checking, model selection, choice of $F$-ratios in the selection of regressors, Hausman's specification error test, tests of non-nested hypotheses (Chapter 12).

These are not new topics for advanced texts but these topics (most of which are never dealt with in introductory texts) are important in econometric practice. The book explains them all with simple models so that students (and researchers) who have not had exposure to advanced texts and advanced courses in econometrics can still learn them and use them.

I have avoided any special discussion of computer programs. Nowadays, there are several computer packages available that one can choose from. Also, they change so rapidly that the lists have to be expanded and updated very often. The study guide will provide some guidance on this. I feel that it is more important to know the answers to the questions of what to compute, why to compute, how to interpret the results, what is wrong with the model, and how we can improve the model. Learning how to compute is rather easy. In the 1960's this last question received more attention because computer technology had not progressed enough and not many efficient computer programs were readily available. With the advances in computer technology and the large number of computer programs readily available, everyone can easily learn "how to compute." That is why I have tried to minimize the discussion of computer programs or the presentation of computer output. Moreover, there is no single program that will do all the things discussed in the book. But by simple adaptation many of the computer programs available will do the job.

Instructors using the book might find it difficult to cover all the material. However, it is always better to have some topics in the book that instructors can choose from depending on their interests and needs. Chapters 1 through 9 cover the basic material. The last three chapters, on models of expectations, errors in variables, and model selection, are topics that instructors can pick and choose from.

A one semester course would cover Chapters 1 to 6 (or 7). A two-semester course would cover Chapters 1 to 9 and parts of Chapters 10–12. In either case Chapter 2 need not be covered but can be used as a reference.

At some places I have a reference to my earlier book: *Econometrics* (New York, McGraw-Hill, 1977) for details. I saw no point in reproducing some proofs or derivations when they were not absolutely necessary for understanding the points being made. At several others, there is a footnote saying that the result or proof can be found in many books in econometrics, and I give a reference to my earlier book with page numbers. However, the same result can be often found in other standard references, such as:

J. Johnston, *Econometric Methods* (New York, McGraw-Hill), 3rd ed., 1984.
J. Kmenta, *Elements of Econometrics* (New York, The Macmillan Co.), 2nd ed., 1986.
H. Theil, *Principles of Econometrics* (New York, Wiley), 1971.
G. G. Judge, C. R. Hill, W. E. Griffiths, H. Lütkepohl and T. C. Lee, *Theory and Practice of Economics* (New York, Wiley), 2nd ed., 1985.
E. Malinvaud, "Statistical Methods of Econometrics" (Amsterdam, North Holland), 3rd ed., 1976.

Since I did not find it practicable to give detailed page numbers for every book, I have just referred to my earlier book.

I would like to thank Kajal Lahiri, Scott Shonkwiler, and Robert P. Trost for going through the whole manuscript and offering helpful comments.

**G. S. M.**

# Contents

# 3

## SIMPLE REGRESSION 27



# 4

## MULTIPLE REGRESSION 89

# 5

## HETEROSKEDASTICITY 159



# 6

## AUTOCORRELATION 185

# 7

## MULTICOLLINEARITY 223



# 8

## DUMMY VARIABLES AND TRUNCATED VARIABLES 251

# 9

## SIMULTANEOUS EQUATIONS MODELS 293



# 10

## MODELS OF EXPECTATIONS 337

# 1

# What Is Econometrics?



## 1.1 WHAT IS ECONOMETRICS?

Literally speaking, the word "econometrics" means "measurement in economics." This is too broad a definition to be of any use because most of economics is concerned with measurement. We measure our gross national product, employment, money supply, exports, imports, price indexes, and so on. What we mean by *econometrics* is:

> The application of statistical and mathematical methods to the analysis of economic data, with a purpose of giving empirical content to economic theories and verifying them or refuting them.

In this respect econometrics is distinguished from mathematical economics, which consists of the application of mathematics only, and the theories derived need not necessarily have an empirical content.

The application of statistical tools to economic data has a very long history. Stigler[1] notes that the first "empirical" demand schedule was published in 1699 by Charles Davenant and that the first modern statistical demand studies were made by Rodulfo Enini, an Italian statistician, in 1907. The main impetus to the development

[1] G. J. Stigler, "The Early History of Empirical Studies of Consumer Behavior, "*The Journal of Political Economy*, 1954 [reprinted in G. J. Stigler, *Essays in the History of Economics* (Chicago: University of Chicago Press, 1965)].

of econometrics, however, came with the establishment of the Econometric Society in 1930 and the publication of the journal *Econometrica* in January 1933.

Before any statistical analysis with economic data can be done, one needs a clear mathematical formulation of the relevant economic theory. To take a very simple example, saying that the demand curve is downward sloping is not enough. We have to write the statement in mathematical form. This can be done in several ways. For instance, defining $q$ as the quantity demanded and $p$ as price, we can write

$$q = \alpha + \beta p \qquad \beta < 0$$

or

$$q = Ap^{\beta} \qquad \beta < 0$$

As we will see later in the book, one of the major problems we face is the fact that economic theory is rarely informative about functional forms. We have to use statistical methods to choose the functional form, as well.

## 1.2 ECONOMIC AND ECONOMETRIC MODELS

The first task an econometrician faces is that of formulating an econometric model. What is a model?

A *model* is a simplified representation of a real-world process. For instance, saying that the quantity demanded of oranges depends on the price of oranges is a simplified representation because there are a host of other variables that one can think of that determine the demand for oranges. For instance, income of consumers, an increase in diet consciousness ("drinking coffee causes cancer, so you better switch to orange juice," etc.), an increase or decrease in the price of apples, and so on. However, there is no end to this stream of other variables. In a remote sense even the price of gasoline can affect the demand for oranges.

Many scientists have argued in favor of simplicity because simple models are easier to understand, communicate, and test empirically with data. This is the position of Karl Popper[2] and Milton Friedman.[3] The choice of a simple model to explain complex real-world phenomena leads to two criticisms:

1. The model is oversimplified.
2. The assumptions are unrealistic.

For instance, in our example of the demand for oranges, to say that it depends on only the price of oranges is an oversimplification and also an unrealistic assumption. To the criticism of oversimplification, one can argue that it is better to start with a simplified model and progressively construct more complicated models. This is the

[2] K. F. Popper, *The Logic of Scientific Discovery* (London: Hutchinson, 1959), p. 142.
[3] M. Friedman, "The Methodology of Positive Economics," in *Essays in Positive Economics* (Chicago: University of Chicago Press, 1953), p. 14.

idea expressed by Koopmans.[4] On the other hand, there are some who argue in favor of starting with a very general model and simplifying it progressively based on the data available.[5] The famous statistician L. J. (Jimmy) Savage used to say that "a model should be as big as an elephant." Whatever the relative merits of this alternative approach are, we will start with simple models and progressively build more complicated models.

The other criticism we have mentioned is that of "unrealistic assumptions." To this criticism Friedman argued that the assumptions of a theory are never descriptively realistic. He says[6]:

> The relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic" for they never are, but whether they are sufficiently good approximations for the purpose at hand. And this question can be answered by only seeing whether the theory works, which means whether it yields sufficiently accurate predictions.

Returning to our example of demand for oranges, to say that it depends only on the price of oranges is a descriptively unrealistic assumption. However, the inclusion of other variables, such as income and price of apples in the model, does not render the model more descriptively realistic. Even this model can be considered to be based on unrealistic assumptions because it leaves out many other variables (like health consciousness, etc.). But the issue is which model is more useful for predicting the demand for oranges. This issue can be decided only from the data we have and the data we can get.

In practice, we include in our model all the variables that we think are relevant for our purpose and dump the rest of the variables in a basket called "disturbance." This brings us to the distinction between an economic model and an econometric model.

An *economic model* is a set of assumptions that approximately describes the behavior of an economy (or a sector of an economy). An *econometric model* consists of the following:

1. A set of behavioral equations derived from the economic model. These equations involve some observed variables and some "disturbances" (which are a catchall for all the variables considered as irrelevant for the purpose of this model as well as all unforeseen events).
2. A statement of whether there are errors of observation in the observed variables.
3. A specification of the probability distribution of the "disturbances" (and errors of measurement).

With these specifications we can proceed to test the empirical validity of the economic model and use it to make forecasts or use it in policy analysis.

[4] T. C. Koopmans, *Three Essays on the State of Economics Science* (New York: McGraw-Hill, 1957), pp. 142–143.

[5] This is the approach suggested by J. D. Sargan and notably David F. Hendry.

[6] Friedman, "Methodology," pp. 14–15.

Taking the simplest example of a demand model, the econometric model usually consists of:

1. The behavioral equation

$$q = \alpha + \beta p + u$$

where $q$ is the quantity demanded and $p$ the price. Here $p$ and $q$ are the observed variables and $u$ is a disturbance term.

2. A specification of the probability distribution of $u$ which says that $E(u \mid p) = 0$ and that the values of $u$ for the different observations are independently and normally distributed with mean zero and variance $\sigma^2$.

With these specifications one proceeds to test empirically the law of demand or the hypothesis that $\beta < 0$. One can also use the estimated demand function for prediction and policy purposes.

## 1.3 THE AIMS AND METHODOLOGY OF ECONOMETRICS

The aims of econometrics are:

1. Formulation of econometric models, that is, formulation of economic models in an empirically testable form. Usually, there are several ways of formulating the econometric model from an economic model because we have to choose the functional form, the specification of the stochastic structure of the variables, and so on. This part constitutes the *specification aspect* of the econometric work.
2. Estimation and testing of these models with observed data. This part constitutes the *inference aspect* of the econometric work.
3. Use of these models for prediction and policy purposes.

During the 1950s and 1960s the inference aspect received a lot of attention and the specification aspect very little. The major preoccupation of econometricians had been the statistical estimation of correctly specified econometric models. During the late 1940s the Cowles Foundation provided a major breakthrough in this respect, but the statistical analysis presented formidable computational problems. Thus the 1950s and 1960s were spent mostly in devising alternative estimation methods and alternative computer algorithms. Not much attention was paid to errors in the specification or to errors in observations.[7] With the advent of high-speed computers, all this has, however, changed. The estimation problems are no longer formidable and econometricians have turned attention to other aspects of econometric analysis.

[7] There was some work on specification errors in the early 1960s by Theil and Griliches, but this referred to omitted-variable bias (see Chapter 4). Griliches' lecture notes (unpublished) at the University of Chicago were titled "Specification Errors in Econometrics."

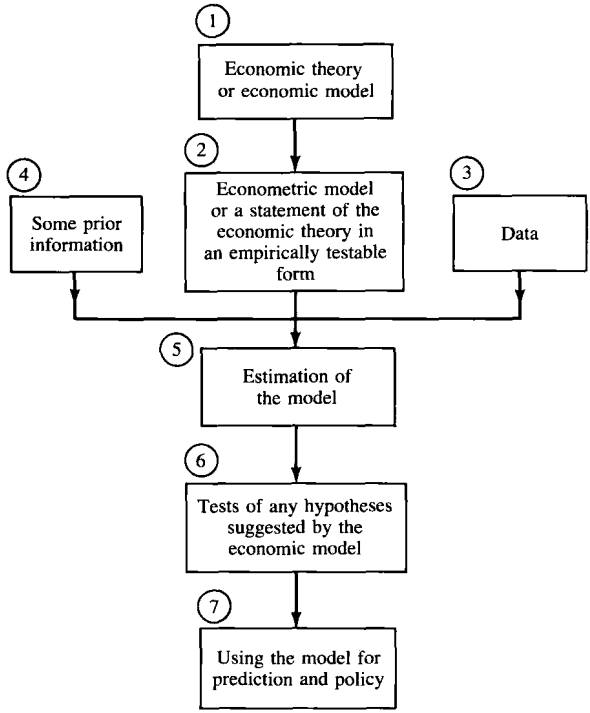



**Figure 1.1.** Schematic description of the steps involved in an econometric analysis of economic models.

We can schematically depict the various steps involved in an econometric analysis, as was done before the emphasis on specification analysis. This is shown in Figure 1.1. Since the entries in the boxes are self-explanatory, we will not elaborate on them. The only box that needs an explanation is box 4, "prior information." This refers to any information that we might have on the unknown parameters in the model. This information can come from economic theory or from previous empirical studies.

There has, however, been considerable dissatisfaction with the scheme shown in Figure 1.1. Although one can find instances of dissatisfaction earlier, it was primarily during the 1970s that arguments were levied against the one-way traffic shown in Figure 1.1. We will discuss three of these arguments.

1. In Figure 1.1 there is no feedback from the econometric testing of economic theories to the formulation of economic theories (i.e., from box 6 to box 1). It has been argued that econometricians are not just handmaidens of economic theorists. It is not true that they just take the theories they are given them and test them, learning nothing from the tests. So we need an arrow from box 6 to box 1.