
BASIC ECONOMETRICS

Damodar Gujarati

BASIC ECONOMETRICS

Damodar Gujarati

*Bernard Baruch College
City University of New York*

INTERNATIONAL STUDENT EDITION

McGRAW-HILL KOGAKUSHA, LTD.

Tokyo Auckland Beirut Bogota Düsseldorf Johannesburg
Lisbon London Lucerne Madrid Mexico Delhi Panama
Paris San Juan São Paulo Singapore Sydney

This book was set in Times Roman. The editors were J. S. Dietrich and Frances A. Neal; the cover was designed by John Hite; the production supervisor was Dominick Petrellese. The drawings were done by Fine Line Illustrations, Inc.

BASIC ECONOMETRICS

INTERNATIONAL STUDENT EDITION

Exclusive rights by McGraw-Hill Kogakusha, Ltd. for manufacture and export. This book cannot be re-exported from the country to which it is consigned by McGraw-Hill.

I .

Copyright © 1978 by McGraw-Hill, Inc.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher.

Library of Congress Cataloging in Publication Data

Gujarati, Damodar .

Basic econometrics

Bibliography p

Includes indexes

1st Econometrics I Title

HB139G84 330 01 82 77 11931

ISBN 0-07-025182-7

PREFACE

The primary objective of *Basic Econometrics* is to provide an elementary but comprehensive introduction to the art and science of econometrics, a field which is becoming increasingly an integral part of training in business and economics.

The book assumes no matrix algebra, calculus, or statistics beyond the elementary level, and it is directed to three classes of readers, namely, junior and senior level undergraduates, beginning graduate students in these disciplines who do not have much prior training in econometrics, and researchers in business, industry, and government who are looking for an informal yet systematic treatment of econometric methods.

Basic ideas of econometrics are presented as simply and as directly as possible without recourse to advanced mathematics and mathematical statistics. Mathematical derivations are kept to an absolute minimum and are usually relegated to the appendixes. This is done deliberately so that the beginner can get a feel for the subject without becoming bogged down in mathematical derivations. The basic philosophy of the book is that econometrics, although grounded in higher mathematics and mathematical statistics, can be taught to the beginner in such a manner that he or she can acquire a fairly good intuitive feeling about the field.

There are several distinguishing features of this book. It is written in very simple and understandable language: matrix algebra or calculus is not essential to follow the text. On the principle of "learning by doing," each econometric technique is illustrated by charts, tables, and real-life economic data. Thus, the reader is made aware of the situations in which a particular economic method is appropriate. The practitioner will find Parts II and III of the text especially appealing. These parts discuss what may be called the "bread-and-butter" aspects of econometrics. Topics such as multicollinearity, heteroscedasticity, and autocorrelation, which are frequently encountered in practice, are discussed logically in a standard format, namely: What is the nature of the problem, what are its consequences, how does one find out whether a problem exists, and what can be

done to remedy the problem? This step-by-step approach will help the researcher to solve a problem systematically.

The extensive and varied exercises at the end of each chapter will distinguish this book from many others; most of the currently available textbooks have very few exercises which are based on real economic data. In all there are about 230 exercises. Detailed solutions to all these exercises are given in the *Instructor's Manual*. Answers and hints to solutions of some 60 exercises are given at the end of the text.

I have included a fairly extensive bibliography which lists several advanced books and articles in econometrics. It is hoped that my book will whet the appetite of the reader so that he or she can approach some of the mathematically advanced material given in the advanced textbooks with less apprehension.

This text can be used in several ways as follows:

Courses in Econometrics

A two-semester course: Chapters 1 to 18, Appendixes A and B. (If matrix algebra is not required, Chapter 8 and Appendix B can be omitted.)

A one-semester course: Chapters 1 to 11, Appendixes A and B. (Again Chapter 8 and Appendix B are optional if matrix approach is not used.)

Courses in Regression Analysis

A one-semester course: Chapters 1 to 11, Appendixes A and B. (Chapter 8 and Appendix B are optional if matrixes are not used.)

A short course of one quarter (about 7 to 8 weeks): Chapters 1 to 7 and Appendix A. (If matrix approach is included, add Chapter 8 and Appendix B.)

In writing this book I have received help from several people. My foremost thanks go to Professor Ann R. Horowitz of the University of Florida and my colleague Professor Albert Zucker. Ann Horowitz's searching and very constructive comments have substantially improved the organization and quality of the text. Al Zucker spent countless hours in the reading of the manuscript and making very valuable suggestions both as to the style and substance. My colleagues, Professors Peter Gutmann, Jay Lee, and Vincent Su, made some very useful suggestions on various chapters, and I am grateful to them. I am also indebted to Professor Giles Burgess of the Portland State University, Professor Ramu Ramanathan of the University of California at San Diego, and Professor Bernard J. Marks of Wichita State University, who read the book in manuscript form and made several constructive suggestions to improve the quality of the text. Needless to say, none of these people are responsible for any deficiencies that may remain. I would also like to acknowledge my debt to my students who read the book while it was in preparation and offered constructive suggestions

to improve the substance and form of the topics covered in the text. I am much indebted to my research assistant Joan Rahav for all her help. Finally, my thanks are due to my editor J. Stephen Dietrich and his associate Diane DeLuca for all their behind-the-scene help and encouragement.

Damodar Gujarati

CONTENTS

Preface
Introduction

PART 1 Single-Equation Regression Models

1	The Nature of Regression Analysis	11
1.1	Historical Origin of the Term "Regression"	11
1.2	The Modern Interpretation of Regression	12
1.3	Statistical vs. Functional Dependencies	15
1.4	Regression and Causation	16
1.5	Regression vs. Correlation	16
1.6	Terminology and Notation	17
1.7	The Role of the Computer in Regression Analysis	18
1.8	Summary and Conclusions	19
2	Two-Variable Regression Analysis: Some Basic Ideas	20
2.1	A Hypothetical Example	20
2.2	The Concept of Population Regression Function (PRF)	23
2.3	The Meaning of the Term "Linear"	24
2.4	Stochastic Specification of PRF	25
2.5	The Nature of the Stochastic Disturbance Term	27
2.6	The Sample Regression Function (SRF)	28
2.7	Summary and Conclusions	31
	Exercises	32

3	Two-Variable Regression Model: The Problem of Estimation	36
3.1	The Method of Ordinary Least Squares	36
3.2	Properties of Least-Squares Estimators: The Gauss-Markov Theorem	47
3.3	The Coefficient of Determination r^2 : A Measure of "Goodness of Fit"	47
3.4	An Illustrative Example	50
3.5	Functional Forms of Regression Models	53
3.6	Summary and Conclusions	56
	Exercises	58
	Appendix 3A	63
	3A.1 Derivation of Least-Squares Estimators	63
	3A.2 Variances and Standard Errors of Least-Squares Estimators	63
	3A.3 The Least-Squares Estimator of σ^2	64
	3A.4 Minimum-Variance Property of Least-Squares Estimators (Gauss-Markov Theorem)	65
	3A.5 Computer Printout of the Consumption-Income Example	67
4	Normality Assumption: Classical Normal Linear Regression Model	70
4.1	The Probability Distribution of Disturbances u_i	70
4.2	The Normality Assumption	71
4.3	Properties of OLS Estimators under the Normality Assumption	72
4.4	The Method of Maximum Likelihood (ML)	74
4.5	Summary and Conclusions	74
	Exercises	75
5	Two-Variable Regression: Interval Estimation and Hypothesis Testing	76
5.1	Interval Estimation: Some Basic Ideas	76
5.2	Normal, t , χ^2 , and F Distributions: A Digression	78
5.3	Confidence Intervals for Regression Coefficients β_0 and β_1	79
5.4	Confidence Interval for σ^2	80
5.5	Hypothesis Testing: General Comments	81
5.6	Hypothesis Testing: The Confidence-Interval Approach	82
5.7	Hypothesis Testing: The Test-of-Significance Approach	83
5.8	Regression Analysis and Analysis of Variance	87
5.9	Application of Regression Analysis: The Problem of Prediction	88
5.10	Reporting the Results of Regression Analysis	91
5.11	Summary and Conclusions	92
	Exercises	93
	Appendix 5A	96
	5A.1 Derivation of Equation (5.3.2)	96
	5A.2 Derivation of Equation (5.8.1)	97

6	Multiple Regression Analysis: The Problem of Estimation	98
6.1	The Three-Variable Model: Notation and Assumptions	98
6.2	Interpretation of Multiple Regression Equation	100
6.3	The Meaning of Partial Regression Coefficients	101
6.4	OLS and ML Estimation of the Partial Regression Coefficients	102
6.5	The Multiple Coefficient of Determination R^2 and the Multiple Coefficient of Correlation R	106
6.6	An Illustrative Example: The Cobb-Douglas Production Function	107
6.7	Comparing Two or More R^2 Values: The Adjusted R^2	109
6.8	Partial Correlation Coefficient	111
6.9	Some Interesting Relationships	114
6.10	Summary and Conclusions	115
	Exercises	115
	Appendix 6A	118
6A.1	Derivation of OLS Estimators Given in Equations (6.4.3) to (6.4.5)	118
6A.2	Equivalence between a_1 of (6.3.5) and $\beta_{12.3}$ of (6.4.7)	119
6A.3	Derivation of Equation (6.4.14)	120
6A.4	Computer Printout of the Cobb-Douglas Production Function Fitted to the Taiwanese Agricultural Sector	121
7	Multiple Regression Analysis: The Problem of Inference	123
7.1	The Normality Assumption Once Again	123
7.2	An Illustrative Example	124
7.3	Hypothesis Testing about Individual Partial Regression Coefficients	126
7.4	Testing the Overall Significance of the Sample Regression	128
7.5	The Analysis-of-Variance Approach to Testing the Overall Significance of an Observed Multiple Regression	129
7.6	An Important Relationship between R^2 and F	131
7.7	The "Incremental," or "Marginal," Contribution of an Explanatory Variable	132
7.8	Testing the Significance of Correlation Coefficients	135
7.9	Summary and Conclusions	136
	Exercises	137
	Appendix 7A	140
7A.1	Computer Printout of the Personal Consumption Expenditure and Personal Disposable Income Regression	140
8	The Matrix Approach to Linear Regression Model*	142
8.1	The k -Variable Linear Regression Model	142
	* This is an optional chapter.	

8.2	Assumptions of the Classical Linear Regression Model in Matrix Notation	144
8.3	OLS Estimation	147
8.4	The Coefficient of Determination R^2 in Matrix Notation	151
8.5	Hypothesis Testing in Matrix Notation	152
8.6	Analysis of Variance in Matrix Notation	153
8.7	Matrix Formulation in Deviation Form	154
8.8	The Correlation Matrix	154
8.9	Summary of the Matrix Approach: An Illustrative Example	156
8.10	Summary and Conclusions	160
	Exercises	161
	Appendix 8A	165
	8A.1 Derivations of k Normal or Simultaneous Equations	165
	8A.2 Matrix Derivation of Normal Equations	166
	8A.3 Variance-Covariance Matrix of $\hat{\beta}$	166

PART 2 Violations of the Assumptions of the Classical Model

9	Multicollinearity	171
9.1	The Nature of Multicollinearity	171
9.2	Estimation in the Presence of Perfect Multicollinearity	174
9.3	Estimation in the Presence of "High" but "Imperfect" Multicollinearity	175
9.4	Consequences of Multicollinearity	177
9.5	An Illustrative Example	179
9.6	Detection of Multicollinearity	181
9.7	Multicollinearity and Prediction	183
9.8	Remedial Measures	183
9.9	Summary and Conclusions	187
	Exercises	188
10	Heteroscedasticity	193
10.1	The Nature of Heteroscedasticity	193
10.2	Consequences of Heteroscedasticity	197
10.3	Detection of Heteroscedasticity	200
10.4	Remedial Measures	206
10.5	Summary and Conclusions	211
	Exercises	212
	Appendix 10A	217
	10A.1 The Method of Weighted Least Squares	217
11	Autocorrelation	219
11.1	The Nature of the Problem	219
11.2	Consequences of Autocorrelation	225

11.3	Detecting Autocorrelation	232
11.4	Remedial Measures	239
11.5	An Illustrative Example	242
11.6	Summary and Conclusions	244
	Exercises	245

PART 3 Topics in Econometrics

12	Autoregressive and Distributed-Lag Models	255
12.1	The Role of "Time," or "Lag," in Economics	256
12.2	The Reasons for Lags	258
12.3	Estimation of Distributed-Lag Models	259
12.4	The Koyck Approach to Distributed-Lag Models	261
12.5	Rationalization of the Koyck Model: The Adaptive Expectation Model	263
12.6	Another Rationalization of the Koyck Model: The Stock Adjustment, or Partial Adjustment, Model	265
12.7	Estimation of Autoregressive Models	266
12.8	The Method of Instrumental Variables	268
12.9	Detecting Autocorrelation in Autoregressive Models: Durbin h Test	269
12.10	An Illustrative Example	270
12.11	The Almon Approach to Distributed-Lag Models: The Almon Polynomial Lag	272
12.12	Summary and Conclusions	279
	Exercises	281
13	Regression on Dummy Variables	287
13.1	The Nature of Dummy Variables	287
13.2	Regression on One Quantitative Variable and One Qualitative Variable with Two Classes, or Categories	289
13.3	Regression on One Quantitative Variable and One Qualitative Variable with More than Two Classes	291
13.4	Regression on One Quantitative Variable and Two Qualitative Variables	293
13.5	The Economics of "Moonlighting": An Application	294
13.6	Comparing Two Regressions	295
13.7	Testing the Equivalency of Two Regressions: An Illustration	298
13.8	The Use of Dummy Variables in Seasonal Analysis	300
13.9	Piecewise Linear Regression	302
13.10	Summary and Conclusions	303
	Exercises	304
	Appendix 13A	309
	13A.1 Data Matrix for Regression (13.7.2)	309
	13A.2 Data Matrix for Regression (13.8.2)	311

14	Regression on Dummy Dependent Variable	312
14.1	Dummy Dependent Variable	312
14.2	Estimation of Linear Probability Models	314
14.3	Cohen-Rea-Lerman Study	316
14.4	Summary and Conclusions	318
	Exercises	319
15	Single-Equation Models: Some Further Topics	322
15.1	Errors of Measurement	322
15.2	Linear Equality Restrictions: Restricted Least Squares	325
15.3	Nonlinear (in-the-Parameters) Regressions	326
15.4	Specification Bias	327
	Exercises	328

PART 4 Simultaneous-Equations Models

16	Simultaneous-Equations Models	335
16.1	The Nature of Simultaneous-Equations Models	335
16.2	Examples of Simultaneous-Equations Models	336
16.3	The Simultaneous-Equations Bias: Inconsistency of OLS Estimators	342
16.4	Summary and Conclusions	344
	Exercises	345
17	The Identification Problem	348
17.1	Notations and Definitions	348
17.2	The Identification Problem	352
17.3	Rules for Identification	359
17.4	Summary and Conclusions	365
	Exercises	365
18	Simultaneous-Equations Methods	368
18.1	Approaches to Estimation	368
18.2	Recursive Models and Ordinary Least Squares	370
18.3	Estimation of a Just Identified Equation: The Method of Indirect Least Squares (ILS)	372
18.4	Estimation of an Overidentified Equation: The Method of Two-Stage Least Squares (2SLS)	376
18.5	Summary and Conclusions	381
	Exercises	382
	Appendix 18A	385
	18A.1 Bias in the Indirect Least Squares Estimators	385
	18A.2 Estimation of Standard Errors of 2SLS Estimates	386

Appendixes	389
A A Review of Some Statistical Concepts	389
B Rudiments of Matrix Algebra	409
C A List of Some Packaged Computer Programs	424
D Statistical Tables	427
Table D.1 Areas under the Standardized Normal Distribution	428
Table D.2 Percentage Points of the t Distribution	429
Table D.3 Upper Percentage Points of the F Distribution	430
Table D.4 Upper Percentage Points of the χ^2 Distribution	436
Table D.5 Durbin-Watson d Statistic: Significance Points of d_L and d_U at 0.05 and 0.01 Levels of Significance	438
Table D.6 Critical Values of Runs in the Run Test	440
Answers and Hints to Solutions of Selected Exercises	442
Selected Bibliography	451
Indexes	457
Name Index	
Subject Index	

INTRODUCTION

1 WHAT IS ECONOMETRICS?

Literally interpreted, *econometrics* means “economic measurement.” Although measurement is an important part of econometrics, the scope of econometrics is much broader, as can be seen from the following quotations.

Econometrics, the result of a certain outlook on the role of economics, consists of the application of mathematical statistics to economic data to lend empirical support to the models constructed by mathematical economics and to obtain numerical results.¹

... econometrics may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference.²

Econometrics may be defined as the social science in which the tools of economic theory, mathematics, and statistical inference are applied to the analysis of economic phenomena.³

Econometrics is concerned with the empirical determination of economic laws.⁴

¹ Gerhard Tintner, *Methodology of Mathematical Economics and Econometrics*, The University of Chicago Press, Chicago, 1968, p. 74.

² P. A. Samuelson, T. C. Koopmans, and J. R. N. Stone, “Report of the Evaluative Committee for *Econometrica*,” *Econometrica*, vol. 22, no. 2, April 1954, pp. 141–146.

³ Arthur S. Goldberger, *Econometric Theory*, John Wiley & Sons, Inc., New York, 1964, p. 1.

⁴ H. Theil, *Principles of Econometrics*, John Wiley & Sons, Inc., New York, 1971, p. 1.

2 WHY A SEPARATE DISCIPLINE?

As the preceding definitions suggest, econometrics is an amalgam of economic theory, mathematical economics, economic statistics and mathematical statistics. Yet, it is a subject that deserves to be studied in its own right for the following reasons.

Economic theory makes statements or hypotheses that are mostly qualitative in nature. For example, microeconomic theory states that, other things remaining the same, a reduction in the price of a commodity is expected to increase the quantity demanded of that commodity. Thus, economic theory postulates a negative or inverse relationship between the price and quantity demanded of a commodity. But the theory itself does not provide any numerical measure of the relationship between the two; that is, it does not tell by how much the quantity will go up or down as a result of a certain change in the price of the commodity. It is the job of the econometrician to provide such numerical estimates. Stated differently, it is econometrics that gives empirical content to most economic theory.

The main concern of mathematical economics is to express economic theory in mathematical form (equations) without regard to measurability or empirical verification of the theory. Econometrics, as noted previously, is mainly interested in the empirical verification of economic theory. As we shall see, the econometrician often uses the mathematical equations proposed by the mathematical economist but puts these equations in such a form that they lend themselves to empirical testing. And this conversion of mathematical into econometric equations requires a great deal of ingenuity and practical skill.

Economic statistics is mainly concerned with collecting, processing, and presenting economic data in the form of charts and tables. This is the job of the economic statistician. It is he or she who is primarily responsible for collecting data on GNP, employment, unemployment, prices, etc. The data thus collected constitute the raw data for econometric work. But the economic statistician does not go any further, not being concerned with using the collected data to test economic theories. Of course, one who does that becomes an econometrician.

Although mathematical statistics provides many of the tools used in the trade, the econometrician often needs special methods in view of the unique nature of most economic data, namely, that the data are not generated as the result of a controlled experiment. The econometrician, like the meteorologist, generally depends on data that cannot be controlled directly. Thus, data on consumption, income, investment, savings, prices, etc., which are collected by public and private agencies, are nonexperimental data. The econometrician takes these data as given. This creates special problems not normally dealt with in mathematical statistics. Moreover, such data are likely to contain errors of measurement, and the econometrician may be called upon to develop special methods of analysis to deal with such errors of measurement.

3 METHODOLOGY OF ECONOMETRICS

To illustrate the methodology of econometrics, let us consider the keynesian theory of consumption. Keynes states:

The fundamental psychological law ... is that men [women] are disposed, as a rule and on average, to increase their consumption as their income increases, but not by as much as the increase in their income.⁵

In short, Keynes postulates that the marginal propensity to consume (MPC), the rate of change of consumption for a unit (say, a dollar) change in income, is greater than 0 but less than 1. To test this theory, the econometrician may proceed as follows.

Specification of the Econometric Model

Although Keynes postulates a positive relationship between consumption and income, he does not specify the precise form of the functional relationship between the two. For simplicity, a mathematical economist may suggest the following form for Keynes' consumption function:

$$Y = \alpha + \beta X \quad (1)$$

where Y = consumption expenditure

X = income

α, β = constants or parameters

The slope coefficient β represents the MPC.

Equation (1), which states that consumption is linearly related to income, is an example of a mathematical model. A model is simply a set of mathematical equations. If the model has only one equation, as in the preceding example, it is called a *single-equation model*, whereas if it has more than one equation, it is known as a *multiequation* or *simultaneous-equation model*.

The purely mathematical model of the consumption function given in (1) is, however, of limited interest to the econometrician for it assumes that there is an exact or deterministic relationship between consumption and income. But relationships between economic variables are generally inexact. Thus, if we were to obtain data on consumption expenditure and disposable (after-tax) income of a sample of, say, 5000 American families and plot these data on a graph paper with consumption expenditure on the vertical axis and disposable income on the horizontal axis, we would not expect all 5000 observations to lie exactly on the straight line of equation (1). This is because in addition to income there are other variables which also affect consumption expenditure. For example, size of family, ages of the members in the family, family religion, etc., are likely to exert some influence on consumption.

⁵ John Maynard Keynes, *The General Theory of Employment, Interest and Money*, Harcourt Brace Jovanovich, Inc., New York, 1936, p. 96.

To allow for the inexact relationships between economic variables, the econometrician would modify the deterministic consumption function (1) as follows:

$$Y = \alpha + \beta X + u \quad (2)$$

where u , known as the *disturbance*, or *error*, term, is a random (stochastic) variable which has well-defined probabilistic properties. The disturbance term u may represent all those forces that affect consumption but are not taken into account explicitly.

Equation (2) is an example of an econometric model. More technically, (2) is an example of a linear regression model, which is a major concern of this book. The econometric consumption function (2) hypothesizes that the dependent variable Y (consumption) is linearly related to the explanatory variable X (income), but the relationship between the two is not exact; it is subject to individual variation.

Estimation

Having specified the econometric model, the next task of the econometrician is to obtain estimates (numerical values) of the parameters of the model from the data available; these data may be provided by the economic statistician. These estimates give empirical content to economic theory. Thus, if in a study of the Keynesian consumption function given previously it is found that $\beta = 0.8$, this value not only provides a numerical estimate of MPC, but also supports Keynes' hypothesis that MPC is less than 1.

How does one estimate the parameters, such as α and β ? An answer to this question will be provided in the following chapters. Suffice it to note here that the statistical tool of regression analysis is the main technique used in this book to obtain the estimates.

Verification (Statistical Inference)

Having obtained estimates of the parameters, the next task of the econometrician is to develop suitable criteria to find out whether the estimates obtained are in conformity with the expectations of the theory that is being tested. As noted previously, Keynes expected the MPC to be positive but less than 1. Suppose in a study of the consumption function it is found that the MPC = 0.9. Although numerically this estimate is less than 1, one may enquire whether the estimate is sufficiently below unity to convince us that this is not an accidental outcome of sampling process. In other words, is this estimate statistically less than 1? If it is, it supports Keynes' contention; otherwise it may refute it.

Such confirmation or refutation of economic theories on the basis of empirical evidence is based on a branch of statistical theory known as *statistical inference* (hypothesis testing). Throughout this book we shall see how this inference process is actually conducted.