ECONOMETRIC METHODS

FOURTH EDITION

JACK JOHNSTON
JOHN DiNARDO

3.5" data disk enclosed. Formatted in ASCII and adaptable to most statistical analysis programs.

ECONOMETRIC METHODS

Fourth Edition

Jack Johnston
University of California, Irvine

John DiNardo
University of California, Irvine

The McGraw-Hill Companies, Inc.
New York St. Louis San Francisco Auckland Bogotá Caracas Lisbon London Madrid Mexico City Milan Montreal New Delhi San Juan Singapore Sydney Tokyo Toronto

McGraw-Hill

A Division of The ***McGraw-Hill*** *Companies*

This book was set in Times Roman by Publication Services, Inc.
The editors were Lucille Sutton and Curt Berkowitz;
the production supervisor was Kathy Porzio.
The cover was designed by Rafael Hernandez.
Project supervision was done by Publication Services, Inc.
R.R. Donnelley & Sons was printer and binder.

ECONOMETRIC METHODS

This book is printed on acid-free paper.

234567890 DOCDOC 90987

P/N 032720-3
Part of ISBN 0-07-913121-2

Library of Congress Cataloging-in-Publication Data is available:
LC Card # 96-77116

http://www.mhcollege.com

ABOUT THE AUTHORS

Jack Johnston is Professor Emeritus of Econometrics at the University of California, Irvine. Before coming to Irvine he was the Stanley Jevons Professor of Econometrics at the University of Manchester. He has also taught at the City University of New York, Emory University, Harvard University, Queen's University, Ontario, the University of Wales, and the University of Wisconsin at Madison. He is a Fellow of the Econometric Society. His books are *Statistical Cost Analysis,* (McGraw-Hill, 1960) and the three previous editions of *Econometric Methods* (McGraw-Hill, 1963, 1972, and 1984).

John DiNardo is an Assistant Professor of Economics at the University of California, Irvine, and a Faculty Research Fellow at the National Bureau of Economic Research. Before coming to Irvine he was a Research Associate at RAND. A labor economist and applied econometrician, he has also taught at the Massachusetts Institute of Technology and Princeton University. His most recent work has been published in *Econometrica,* the *Journal of Political Economy,* and the *Quarterly Journal of Economics.* This is his first book.

PREFACE

In the twelve years since the third edition of this book was published, the computing power at the fingertips of econometricians has dramatically increased. Econometric theorists have also been substantially increasing the number of suggested estimation, testing, and diagnostic procedures, most of which were quickly made available in econometric software packages. Faced with this cornucopia, the applied econometrician, whose task is the analysis of real-life data, often suffers from "intellectual indigestion" and finds it difficult to make informed and sensible judgments about which procedures to implement.

In writing this new edition we have had two major objectives. The first is to provide a comprehensive and accessible account of available econometric methods. The second is to illustrate these methods with applications to some real data sets, which are given on the data diskette that accompanies the book; thus, the reader can replicate the applications in the text, experiment with some of the problems suggested at the chapter ends, and carry out further analyses of her own choosing. These objectives have dictated an almost total rewriting of the book and the addition of substantial treatments of new topics that have not appeared in previous editions.

As with earlier editions, it is assumed that the reader has an understanding of the basic concepts of statistical inference. However, Appendix B, on statistics, gives a review of the major topics, and the detailed treatment of inference procedures in the earlier chapters should help with statistical recall. Again, matrix algebra is used extensively in the text. Appendix A, on matrix algebra, provides a comprehensive treatment, where the development matches as far as possible the order in which the various matrix concepts are used in the main text. Thus, a reader new to matrix algebra can switch between the text and Appendix A as the topics require.

Recent econometric developments surveyed in this edition may be grouped into six major areas:

- Asymptotics
- Time series
- Model evaluation

- Generalized method of moments
- Computationally intensive methods
- Microeconometrics

Asymptotics

Realistic model specifications often do not permit the development of exact, finite-sample results. It is, however, frequently possible to derive results that hold asymptotically. The maximum likelihood principle is used extensively in recent work, and the classical likelihood ratio, Wald, and Lagrange multiplier tests are frequently applied. An introduction to asymptotic results and to the maximum likelihood principle is given in Chapter 2 in the context of the two-variable model, where the regressor is the lagged value of the dependent variable. The student is thus introduced to these basic concepts at an early stage. Chapter 5 gives an extended treatment of maximum likelihood and the trinity of classical tests. Chapter 6, on heteroscedasticity and autocorrelation, describes many applications of these tests.

Time Series

Analysis of univariate time series continues to be an important topic, but the major new development in this field is the investigation of nonstationary series and the impact of nonstationarity on estimation procedures. Thus, one needs to test for stationarity, and a large literature has developed around unit root tests. When a regression is run containing two or more nonstationary series, there is the possibility that some linear combination of these series has stationary residuals, in which case the series are said to be *cointegrated.* Tests for the possible existence of cointegration are thus important, and estimation procedures for a given data set depend on the number of cointegrating relations found. The contrast between stationary and nonstationary series is introduced in Chapter 2 in the context of the two-variable model, where the regressor is the lagged value of the dependent variable. This is developed fully in Chapter 7, which is devoted to the analysis of univariate time series. This chapter closes with an empirical application to monthly data on U.S. housing starts. Chapter 8 contains an extensive discussion of cointegration tests and estimation procedures. These are illustrated in an empirical study of gasoline demand.

Model Evaluation

There has been intense debate on model evaluation and diagnostic procedures. The debate continues and there is, as yet, little consensus. However, it does seem that more applied researchers are conducting various evaluation tests. A basic principle of this approach is to divide sample data into two subsets: one to be used for estimating some specified model and the other to be used for evaluating the results of the estimation. Chapter 4 illustrates the application of many of these tests to a least-squares, linear regression model. Chapter 8 contains a detailed account of the use of diagnostic tests in the development of a model of the demand for gasoline.

Generalized Method of Moments (GMM)

Led initially by developments in macroeconomics, in particular "Euler equation approaches," GMM has become an increasingly important topic and has been given its own separate chapter (Chapter 10). As work in this area has developed, it has become apparent that GMM also provides a pedagogically useful way to look at old questions. In particular, the role of the "orthogonality condition" is highlighted as an organizing framework for looking at some old problems (OLS, 2SLS, Hausman tests, and even classical experimental design.)

Computationally Intensive Methods

One consequence of the "computer revolution" is the increased use of methods that, not too many years earlier, were computationally prohibitive. In Chapter 11, we review several of these methods: Monte Carlo methods, the bootstrap, permutation tests, and nonparametric estimation methods. As several of these techniques have merited separate treatises, it is impossible to cover the topics comprehensively. Instead the chapter aims at a more modest goal: to introduce the student to several of these developments and to provide an understanding of some basic principles and their potential range of application. Toward that end, several simple examples are presented in the text in some detail in the hope that the student can begin to use some of these techniques even in more realistic and complex situations.

Microeconometrics

Perhaps nowhere else has the increased sophistication of statistical software made a greater mark on econometric practice than in microeconometric applications. Chapter 12, on panel data, introduces the student to the simplest models—the fixed effect and random effect models—that are routinely applied to the ever-increasing number of panel data sets available. We also attempt to provide the student with practical advice about the advantages and disadvantages of these techniques. In Chapter 13 we review limited dependent variable models. Our review is selective: The literature is so vast, and the techniques available to researchers in statistical programs so numerous, that the temptation to provide a "cookbook" rendition of these topics is very strong. We have resisted the temptation as far as possible. This has meant the omission of some important topics—to name just two, hazard models and models with multiple choices (except the ordered probit). On the other hand, we go through the probit and logit models in some detail using an empirical illustration with the data diskette. As some software packages routinely calculate "Huber" standard errors for the probit and logit, it was felt some discussion of heteroscedasticity in these models and quasi-maximum likelihood was necessary. A discussion of heteroscedasticity in the Tobit led us to include two recent techniques for the censored regression model: "symmetrically trimmed least squares" and "least absolute deviations." The chapter concludes with a brief discussion of the ubiquitous "Heckman correction" and related issues.

Acknowledgments

Many helpful suggestions came from colleagues at the University of California, Irvine, especially Jae-Woo Lee, David Lilien, and Janey Morrison. A special debt is owed to Hiroyuki Kawakatsu, who read the complete manuscript, posed many penetrating and insightful questions, checked all empirical examples, and wrote the Solutions Manual. David Lilien kindly provided the software (MicroTSP and EViews) that has been used extensively in the preparation of the empirical illustrations in the book. Thanks also go to DRI/McGraw-Hill for permission to reproduce from the data set, *DRI Basic Economics,* many of the series on the data diskette. We are also very indebted to Ian Domowitz, Kiseok Lee, and Timothy Vogelsang, who read the entire manuscript and contributed many valuable suggestions.

DiNardo, in particular, would like to thank Julie Berry Cullen, Kristin Butcher, Kenneth Chay, Angus Deaton, Whitney Newey, Jack Porter, his MIT students, and the Econometrics Lunch at MIT for helpful instruction. In addition his warmest thanks go to Cheng Hsiao, Dean Hyslop, Jörn-Steffen Pischke, Gary Solon, Robert Valletta, and Jean Wohlever for extensive comments on earlier drafts of Chapters 10–13, which improved the exposition immensely. He also thanks Dave Card and the Industrial Relations Section at Princeton University, Michael Grossman and the National Bureau of Economic Research in New York, and the Department of Economics at MIT for their hospitality during the completion of the book.

Jack Johnston
John DiNardo

Some Tests of the *k*-Variable Linear Equation for Specification Error

Maximum Likelihood (ML), Generalized Least Squares (GLS), and Instrumental Variable (IV) Estimators 142

Heteroscedasticity and Autocorrelation 162

2.1.1 Constant Growth Curves

Taking first differences of Eq. (2.5) gives

$$\Delta Y_t = \beta + (u_t - u_{t-1})$$

If we ignore the disturbances, the implication of Eq. (2.5) is that the series increases (decreases) by a constant amount each period. For an increasing series ($\beta > 0$), this implies a *decreasing* growth rate, and for a decreasing series ($\beta < 0$), the specification gives an *increasing* decline rate. For series with an underlying *constant* growth rate, whether positive or negative, Eq. (2.5) is then an inappropriate specification. The appropriate specification expresses the *logarithm* of the series as a linear function of time. This result may be seen as follows.

Without disturbances a constant growth series is given by the equation

$$Y_t = Y_0(1 + g)^t \tag{2.6}$$

where $g = (Y_t - Y_{t-1})/Y_{t-1}$ is the constant proportionate rate of growth per period. Taking logs of both sides of Eq. (2.6) gives[1]

$$\ln Y_t = \alpha + \beta t \tag{2.7}$$

where $$\alpha = \ln Y_0 \qquad \text{and} \qquad \beta = \ln(1 + g) \tag{2.8}$$

If one suspects that a series has a constant growth rate, plotting the log of the series against time provides a quick check. If the scatter is approximately linear, Eq. (2.7) can be fitted by least squares, regressing the log of Y against time. The resultant slope coefficient then provides an estimate $\hat{g}$ of the growth rate, namely,

$$b = \ln(1 + \hat{g}) \qquad \text{giving} \qquad \hat{g} = e^b - 1$$

The β coefficient of Eq. (2.7) represents the *continuous* rate of change $\partial \ln Y_t/\partial t$, whereas g represents the *discrete* rate. Formulating a constant growth series in continuous time gives

$$Y_t = Y_0 e^{\beta t} \qquad \text{or} \qquad \ln Y_t = \alpha + \beta t$$

Finally, note that taking first differences of Eq. (2.7) gives

$$\Delta \ln Y_t = \beta = \ln(1 + g) \approx g \tag{2.9}$$

Thus, taking first differences of logs gives the continuous growth rate, which in turn is an approximation to the discrete growth rate. This approximation is only reasonably accurate for small values of g.

2.1.2 Numerical Example

Table 2.1 gives data on bituminous coal output in the United States by decades from 1841 to 1910. Plotting the log of output against time, we find a linear relationship.

---

[1]We use ln to indicate logs to the natural base e.

TABLE 2.1
Bituminous coal output in the United States, 1841–1910

Decade	Average annual output (1,000 net tons), Y	$\ln Y$	t	$t(\ln Y)$
1841–1850	1,837	7.5159	−3	−22.5457
1851–1860	4,868	8.4904	−2	−16.9809
1861–1870	12,411	9.4263	−1	−9.4263
1871–1880	32,617	10.3926	0	0
1881–1890	82,770	11.3238	1	11.3238
1891–1900	148,457	11.9081	2	23.8161
1901–1910	322,958	12.6853	3	38.0558
Sum		71.7424	0	24.2408

So we will fit a constant growth curve and estimate the annual growth rate. Setting the origin for time at the center of the 1870s and taking a unit of time to be 10 years, we obtain the t series shown in the table. From the data in the table

$$a = \frac{\sum \ln Y}{n} = \frac{71.7424}{7} = 10.2489$$

$$b = \frac{\sum t \ln Y}{\sum t^2} = \frac{24.2408}{28} = 0.8657$$

The r^2 for this regression is 0.9945, confirming the linearity of the scatter. The estimated growth rate per decade is obtained from

$$\hat{g} = e^b - 1 = 1.3768$$

Thus the constant growth rate is almost 140 percent per decade. The annual growth rate (agr) is then found from

$$(1 + \text{agr})^{10} = 2.3768$$

which gives agr $= 0.0904$, or just over 9 percent per annum. The equivalent continuous rate is 0.0866.

The time variable may be treated as a fixed regressor, and so the inference procedures of Chapter 1 are applicable to equations like (2.5) and (2.7).[2]

2.2 TRANSFORMATIONS OF VARIABLES

The log transformation of the dependent variable in growth studies leads naturally to the consideration of other transformations. These transformations may be of the

[2] For a very useful discussion of the use of time as a regressor, see Russell Davidson and James G. MacKinnon, *Estimation and Inference in Econometrics*, Oxford University Press, 1993, pp. 115–118.

dependent variable, the regressor variable, or both. Their main purpose is to achieve a **linearizing transformation** so that the simple techniques of Chapter 1 may be applied to suitably transformed variables and thus obviate the need to fit more complicated relations.

2.2.1 Log-Log Transformations

The growth equation has employed a transformation of the dependent variable. Many important econometric applications involve the logs of both variables. The relevant functional specification is

$$Y = AX^{\beta} \qquad \text{or} \qquad \ln Y = \alpha + \beta \ln X \tag{2.10}$$

where $\alpha = \ln A$. The **elasticity** of Y with respect to X is defined as

$$\text{Elasticity} = \frac{dY}{dX}\frac{X}{Y}$$

It measures the percent change in Y for a 1 percent change in X. Applying the elasticity formula to the first expression in Eq. (2.10) shows that the elasticity of this function is simply β, and the second expression in Eq. (2.10) shows that the slope of the log-log specification is the elasticity. Thus Eq. (2.10) specifies a **constant elasticity function.** Such specifications frequently appear in applied work, possibly because of their simplicity and ease of interpretation, since slopes in log-log regressions are direct estimates of (constant) elasticities. Figure 2.1 shows some typical shapes in the Y,X plane for various βs.

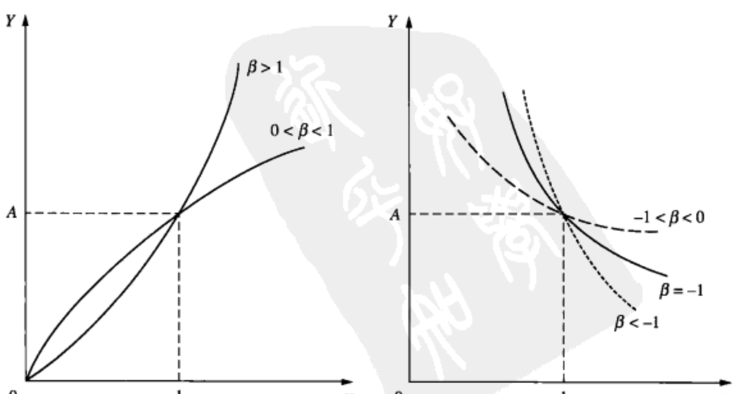

FIGURE 2.1
$Y = AX^{\beta}$.