J. Johnston

ECONOMETRIC METHODS

Third Edition

ECONOMETRIC METHODS

Third Edition

J. Johnston

University of California, Irvine

McGraw-Hill Book Company

New York St. Louis San Francisco Auckland Bogotá Hamburg
Johannesburg London Madrid Mexico Montreal New Delhi
Panama Paris São Paulo Singapore Sydney Tokyo Toronto

IN MEMORY OF
B. and J.

This book was set in Times Roman by Science Typographers, Inc.
The editors were Patricia A. Mitchell and Scott Amerman;
the cover was designed by Nadja Furlan;
the production supervisor was Phil Galea.
The drawings were done by Burmar.
Halliday Lithograph Corporation was printer and binder.

ECONOMETRIC METHODS

1 2 3 4 5 6 7 8 9 0 HALHAL 8 9 8 7 6 5 4

ISBN 0-07-032685-1

Library of Congress Cataloging in Publication Data

Johnston, J. (John), date
Econometric methods.

Includes bibliographical references and index.
1. Econometrics. I. Title.
HB139.J65 1984 330'.028 83-14899
ISBN 0-07-032685-1

PREFACE

This edition has been completely rewritten. The main features of the new edition are the following:

1. The mathematical, statistical threshold has been lowered in order to make the material more accessible to students with only an elementary prior knowledge of statistics. This has resulted in a somewhat larger proportion of words to symbols in the early chapters than would otherwise have been the case. A series of paragraphs on mathematical, statistical topics has also been provided in Appendix A. These are keyed into the early chapters to ease the transition into the heart of the book. For the same reason the chapter on matrix algebra has been retained and, indeed, expanded to include a geometric as well as an algebraic treatment of some topics.
2. All the inference procedures for the general linear model have been derived as special cases of a single basic procedure, namely, the testing of a set of linear restrictions on the parameters of the model (Chapter 5). This leads in turn to an exhaustive treatment of tests for structural change (Chapter 6). Chapter 6 also contains extended treatments of the use of dummy variables and of multicollinearity among the regressors.
3. Every effort has been made to cover both new and old topics on which substantial work has been done in recent years and which are thought to be significant and enduring rather than passing fancies. Such topics include the estimation of sets of equations with special reference to transcendental logarithmic approximations and applications in energy economics (Chapter 8), autocorrelated error terms (Chapter 8), time series techniques (Chapter 9) and, in *A Smorgasbord of Further Topics* (Chapter 10), the following "menu": recursive residuals, spline functions, pooling of time-series and cross-section data, variable-parameter models, qualitative dependent variables, and errors in variables. The author has also granted himself the indulgence of some personal comments on the present state of econometrics (Chapter 12).

4. The problem sets have been extended to become truly Anglo-American, with offerings from the Royal Statistical Society plus the universities of Cambridge, London, Manchester, and Oxford on one side of the pond, and Chicago, Michigan, Yale, and Washington on the other side. Grateful acknowledgment is made to various anonymous authorities for the first set, and to Arnold Zellner, Jan Kmenta, Peter Phillips, and Charles Nelson for the second. Appendix B also contains an extensive set of statistical, econometric tables, and grateful acknowledgment is made to the appropriate sources for permission to publish them.

My debts to many individuals can be warmly acknowledged but never fully recompensed. Craig Riddell (University of British Columbia) read the entire manuscript and contributed many valuable comments. I am very grateful to both of them. My thanks also go to Ian McAvinchey (University of Aberdeen) and to my colleagues Ken Chomitz, Max Fry, and Charles Lave (University of California, Irvine) for a similar service. Ken Chomitz has also produced a solutions manual for comments on various chapters. This in some places is almost a supplementary text in that extended solutions have been written for various problems, outlining particular issues that could not be dealt with in the main text. Copies of the solutions manual are available to instructors on application to the publishers. Kathy Alberti and Barbara Sawyer did a magnificent job on a difficult manuscript. In addition, Barbara Sawyer did the preliminary artwork, prepared the tables in Appendix B, and proofread the entire manuscript. Finally, I gratefully acknowledge the questions and suggestions from teachers and students in many parts of the world during the two decades since the first publication of this book. I can only hope that this third edition will inspire a similar response.

J. Johnston

CONTENTS

CHAPTER
ONE

THE NATURE OF ECONOMETRICS

Before asking the question, "What is econometrics?," one must pose the prior question, "What is economics?" The answer to the second question will indicate the role that econometrics can play in the development of economics. Although the focus of the exposition in this chapter will be on economic models, the methods that have been developed in econometrics can and do play an important role in other social sciences, where there is a concern with building and estimating models of the interconnections between various sets of variables in a predominantly nonexperimental situation.

1-1 ECONOMIC MODEL BUILDING

Economists seek to understand the nature and functioning of economic systems. Their concerns may relate to global aggregates, or *macro* quantities, such as the value of the gross national product (GNP), the level of employment, or the current level of the consumer price index. Alternatively, the focus of attention may be some sector or area of the economy, such as production and employment in the automobile industry or the price and volume of the peanut crop in Georgia. One objective of such an understanding is to be able to make conditional predictions of the likely future development of the system and hopefully enable economic agents, whether government, business, or consumers, to take action to control to some degree the evolution of the system. Another important objective is to test economic theories about the system.

The first step in seeking to understand the functioning of a system is to build a theoretical model. All models are inevitably simplifications of reality, and the model builder seeks to capture the fundamental features of the system being studied. The performance of an economy, or a sector of an economy, at any point in time will depend upon the decisions of various economic agents, taken in the context of the existing state of technology with given stocks of capital, labor, and other limited productive resources. Thus theoretical models typically contain *behavioral relations*, which describe the forces thought to determine the behavior of various groups of economic agents, and *technological relations*, which describe the restrictions imposed by the current technology and endowments of the system. Often technological relations, such as the production function, describing the maximum output achievable with various inputs of capital, labor, and other productive resources, may not appear explicitly in the model, but will have been used in the derivation of behavioral relations, such as the demand function for labor, and so on. In addition to behavioral and technological relations, economic models typically contain *identities* or *definitional relations*.

1-2 A NATIONAL INCOME MODEL

As an example of the model building process let us consider one of the simplest forms of the national income model, which is used as a pedagogic device in most elementary textbooks on economics. Such models begin with the national income identity. For a closed economy with no foreign trade, this identity in any period is

$$y \equiv c + i + g \tag{1-1}$$

where y = gross national product (GNP)
c = consumption expenditure
i = investment expenditure
g = government expenditure

all expenditure flows being measured in real terms. The construction of the model proceeds with the formulation of hypotheses about the determinants of the expenditure components of GNP.

Consumption expenditure might be hypothesized as dependent on disposable income, net of tax, and the rate of interest. Thus we write†

$$c = f((1 - \tau)y, r) \tag{1-2}$$

where τ = tax rate (assumed constant across the economy)
r = rate of interest

The theoretical expectations about this relation are

$$0 < f_1 < 1, \qquad f_2 < 0 \tag{1-3}$$

where f_i indicates the partial derivative of the function with respect to the ith

† See App. A-1, Functions and Derivatives.

argument. The first assumption in Eq. (1-3) is that the marginal propensity to consume out of disposable income is a positive fraction less than unity. The second assumption is that a rise in the rate of interest will have a depressing effect on consumption since it raises the return on savings, increases the cost of financing consumer durables, and also reduces the nominal value of bonds, which are a part of wealth, which in turn might appear as an argument of the consumption function but has been omitted from Eq. (1-2) on grounds of simplicity.

The investment function may be specified as

$$i = f(\Delta y, r) \tag{1-4}$$

with

$$f_1 > 0, \qquad f_2 < 0 \tag{1-5}$$

The term Δy indicates the change in GNP. Investment is positively influenced by profit expectations, and the crude assumption here is that observed changes in real GNP serve as a proxy for these profit expectations. The rate of interest is again expected to be negatively related to this form of expenditure.

Collecting results, so far we have a three-equation model, namely,

$$y \equiv c + i + g$$
$$c = f((1 - \tau)y, r)$$
$$i = f(\Delta y, r)$$

supplemented by the expected signs on derivatives expressed in Eqs. (1-3) and (1-5). This model then constitutes a theory about the joint determination, or "explanation," of the three variables c, i, and y. Such an explanation is obviously *conditional* on the values assumed for g, r, and τ. The model builder now faces a decision on how to treat these remaining variables. Should one formulate theories to explain the determination of government expenditure, the rate of interest, and the tax rate, thus expanding the system to one of six equations? If one does, the new equations will almost certainly contain some explanatory variables on the right-hand side that have not previously appeared in the system, and these, in turn, raise the question of how they are to be treated. It might seem that economic models must become infinitely large, but there is not, of course, an infinite number of variables to be explained. In any case the behavior of model builders is very pragmatic. Everything is relative: all depends on the problem at hand. For some purposes a small model is sufficient and some variables, which in larger models would have explanatory equations, may be left "unexplained." In the present instance we make no pretense at economic realism, but only require a model for illustrative purposes, so we will restrict it to the three equations already specified.

The model contains only two behavioral relations, one for consumption and the other for investment. Economic theory has done two things. First, it has specified the list of explanatory variables on the right-hand side of each equation, and second, it has indicated the expected signs on the partial derivatives. This is usually as far as theory per se can go, but it still leaves a series of important questions unanswered.

1-3 UNANSWERED QUESTIONS

Functional form Theoretical considerations alone cannot usually specify the functional form connecting the variables in a relationship. Many functional forms are consistent with a priori signs on derivatives. Letting

$$z = (1 - \tau)y$$

denote disposable income and omitting the rate of interest variable, the following functional forms all give c as a monotonically increasing function of z and, with appropriate restrictions on parameters, could satisfy the condition that the marginal propensity to consume is a positive fraction:

$$c = \alpha_0 + \alpha_1 z$$

$$c = Az^{\alpha_1}$$

$$c = \alpha_0 - \alpha_1 z^{-1}$$

These functions, however, have different qualitative implications. In the first, an extra \$100 of income always produces the same absolute increase in consumption expenditure. The second and third functions both exhibit a declining marginal propensity to consume as income rises. However, the second function implies that consumption rises indefinitely with income, while the third shows consumption approaching a saturation or asymptotic level α_0 as income becomes very large. This is a typical example of the fact that the qualitative restrictions deriving from economic theory do not serve to delimit functional forms very closely.

Data definition and measurement Theory is sometimes precise and sometimes sloppy in the matter of definitions. In this model, for instance, should consumption be taken to mean expenditure, including actual expenditure on consumer durables, or should consumption of durables be treated as an implicit flow measured by the value of services from the existing stock of consumer durables? If the second definition is taken, is this consistent with the definition of the same variable in the national income identity? What is meant by income? Should it be adjusted for purely seasonal fluctuations or not? Is it to be taken as some recently observed level, or should it be interpreted as some kind of "permanent" or "long run" income? There are many different rates of interest. Should we select a "representative" rate or some combination of rates, and should this variable be treated the same way in both consumption and investment functions?

Lag structure Somewhat allied with problems of data definition are problems of lag structure. Should investment be specified as responding to the current interest rate or to some set of previous interest rates in view of the inevitable time lags involved in making and implementing investment decisions? Again, by the nature of things, economic theory cannot be specific about appropriate lag structures. Moreover, much of economic theorizing has necessarily been about *equilibrium* positions, as, for example, the equilibrium rate of consumption corresponding to some level of income, which has, in theory, remained constant long enough for consumers to become fully adjusted to it. In practice, the world is always staggering from one disequilibrium position to another, so actual data reflect adjustment processes rather than equilibrium positions. Equilibrium theory, by definition, says nothing about adjustment processes, and theories of adaptation and adjustment are still in a fairly primitive state.

Qualitative versus quantitative implications The theoretical model does yield unambiguous qualitative implications, such as that a rise in the rate of interest will depress GNP and a rise in government expenditures will increase it. In more complicated models qualitative conditions on the various equations may not lead to unambiguous predictions about the overall behavior of the model. If our simple model asserted that the rate of interest had a positive effect on consumption and a negative effect on investment, the direction of the rate of interest effect on GNP could not then be known without *quantitative* knowledge of the two separate effects and the magnitudes of consumption and investment. In practice, of course, policymakers are vitally concerned with the likely *magnitude* and *timing* of the effects of changes in the rate of interest, tax rates, or government expenditure. The expected signs of partial derivatives cannot provide this kind of information.

Choice between theories So far, in discussing the previous four problems, we have implicitly assumed that our theoretical model is "correct," but how can we tell whether a theory is sufficiently correct to be used as a valid tool of analysis? Perhaps there are as many theories as there are theorists. There is, in practice, a very important and very difficult problem involved in attempting to discriminate between competing theories. Some theoretical models differ in degree but not in kind. They might be regarded as variations on a theme. For example, another theorist might accept the general form of our consumption and investment functions but wish to add wealth as an additional explanatory variable to the first equation and capital stock to the second. At the other end of the spectrum would be a theorist who rejected the Keynesian flavor of our model and advanced instead a supply-determined theory of output or a model in which the fundamental driving force was the money supply.

1-4 ROLE OF ECONOMETRICS

Econometrics tackles all five questions. Its basic task is to put *empirical* flesh and blood on theoretical structures. This involves several crucial steps. First of all, the theory or model must be specified in explicit functional form. The econometrician does not have any special insights in this area that are denied to the economic theorist, so one usually starts with the simplest functional forms that are consistent with the a priori specifications. At the same time one makes an initial specification of the lag structure. As an example we might specify the three-equation national income model as

$$c_t = \alpha_0 + \alpha_1(1 - \tau)y_t + \alpha_2 r_t \tag{1-6}$$

$$i_t = \beta_0 + \beta_1(y_{t-1} - y_{t-2}) + \beta_2 r_{t-1} \tag{1-7}$$

$$y_t \equiv c_t + i_t + g_t \tag{1-8}$$

with a priori expectations

$$0 < \alpha_1 < 1, \qquad \alpha_2 < 0, \qquad \beta_1 > 0, \qquad \beta_2 < 0$$

The subscripts on the variables refer to time periods. The unit time period can be anything considered relevant by the econometrician, provided there exist appropriate data in terms of that unit. However, it is typically a quarter or a year, and the model is in discrete, not continuous, time.

The second task of the econometrician is to decide on the appropriate data definitions and assemble the relevant data series for the variables which enter the model. The third task is to perform a "marriage" of theory and data by means of statistical methods. The "offspring" of the marriage are various sets of statistics, which shed crucial light on the validity of the theoretical model that has been specified. The most important set consists of the numerical estimates of the parameters of the structural form. The Greek letters of Eqs. (1-6) and (1-7) are now replaced by numbers. There are further statistics which enable one to assess the reliability or precision with which these parameters have been estimated, which in turn helps us to check whether the model conforms to the theoretical expectations about signs of derivatives. There are still further statistics and diagnostic tests that help one to assess the performance of the model and decide whether or not to proceed sequentially by modifying the specification in certain directions and testing out the new variant of the model against the data.

Most of this book will be concerned with the statistical methods used by econometricians in estimating, testing, and evaluating economic models. Historically, econometrics started with the corpus of methods inherited from classical statistics. These methods, however, were mainly developed in the context of the experimental sciences. Special problems of statistical inference arise in economics, where the possibility of controlled experiments is the exception, not the rule, and these will be described in the chapters to follow. All that remains to be done in this introductory chapter is to indicate some of the possible applications of an econometric model, once it has been estimated. This will again be done with the simple model outlined above.

1-5 STRUCTURAL AND REDUCED FORMS

Equations (1-6) to (1-8) constitute the *structural* form of the model. The structural form may be regarded as a theoretical explanation, or hypothesis, about the determination of the three variables y_t, c_t, and i_t, *conditional* on the values currently assumed by g_t and r_t and also on the recent history of the system as represented by y_{t-1}, y_{t-2}, and r_{t-1}. This enables us to make the following classification of the variables in the system:

Current endogenous variables:	c_t, i_t, y_t
Lagged endogenous variables:	y_{t-1}, y_{t-2}
Current exogenous variables:	g_t, r_t
Lagged exogenous variables:	r_{t-1}

The crucial distinction is between endogenous and exogenous variables. The former are those variables whose current values are, in theory, explained by the functioning of the model. The model, however, has nothing to say about the determination of the exogenous variables. A second important distinction is that between the current time period t and previous periods, such as $t - 1$, $t - 2$, and so on. When we come to study the functioning of the model in period t, all lagged values, whether of endogenous or exogenous variables, are already given and cannot now assume new values. Once values are also fed in for the *current* exogenous variables g_t and r_t, the model then delivers the values of the current endogenous variables c_t, i_t, and y_t. This point may be expressed formally by recasting Eqs. (1-6) to (1-8) in an alternative form. Substituting Eqs. (1-6) and (1-7) in Eq. (1-8) and rearranging gives

$$y_t = (\alpha_0 + \beta_0)\delta + \alpha_2\delta r_t + \beta_1\delta(y_{t-1} - y_{t-2}) + \beta_2\delta r_{t-1} + \delta g_t \qquad (1\text{-}9)$$

where

$$\delta = \frac{1}{1 - \alpha_1(1 - \tau)}$$

The important point about Eq. (1-9) is that only one current endogenous variable appears in the equation, namely, y_t on the left-hand side. The right-hand-side variables are a mixture of current exogenous variables and lagged variables, whether endogenous or exogenous. This collection of three sets of variables is labeled the class of *predetermined variables*, since, from the viewpoint of the model in period t, their values either are determined by the past history of the system or are set exogenously in the current period. The investment equation already has nothing but predetermined variables on the right-hand side, so we repeat it here:

$$i_t = \beta_0 + \beta_1(y_{t-1} - y_{t-2}) + \beta_2 r_{t-1} \qquad (1\text{-}10)$$

Finally, substituting Eq. (1-9) in the consumption function gives

$$c_t = [\alpha_0 + \alpha_1(1 - \tau)(\alpha_0 + \beta_0)\delta] + [\alpha_2 + \alpha_1(1 - \tau)\alpha_2\delta]r_t$$
$$+ \alpha_1(1 - \tau)\beta_1\delta(y_{t-1} - y_{t-2}) + \alpha_1(1 - \tau)\beta_2\delta r_{t-1} + \alpha_1(1 - \tau)\delta g_t \qquad (1\text{-}11)$$

The three Eqs. (1-9), (1-10), and (1-11) constitute the *reduced form* of the model. Each equation of the reduced form expresses a current endogenous variable as a function only of predetermined variables. The reduced form may be

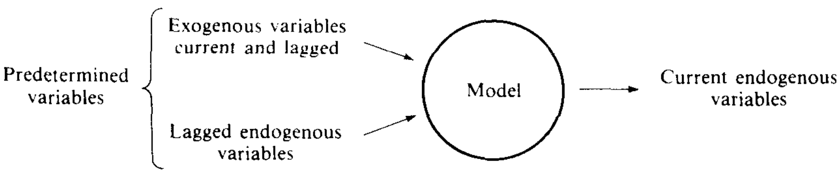

Figure 1-1

written compactly as

$$y_t = \pi_{10} + \pi_{11}g_t + \pi_{12}r_t + \pi_{13}r_{t-1} + \pi_{14}y_{t-1} + \pi_{15}y_{t-2} \tag{1-12}$$

$$c_t = \pi_{20} + \pi_{21}g_t + \pi_{22}r_t + \pi_{23}r_{t-1} + \pi_{24}y_{t-1} + \pi_{25}y_{t-2} \tag{1-13}$$

$$i_t = \pi_{30} + \pi_{33}r_{t-1} + \pi_{34}y_{t-1} + \pi_{35}y_{t-2} \tag{1-14}$$

where the π's are the functions of the structural parameters indicated in Eqs. (1-9) to (1-11). Schematically, the reduced form is indicated in Fig. 1-1.

The reduced form also indicates that there is one-way causation in the model in the sense that the exogenous variables influence the current endogenous variables, but there is no feedback in the opposite direction: current endogenous variables do not influence the exogenous variables.

1-6 MULTIPLIERS AND DYNAMIC PROPERTIES

The π's of the reduced-form equations are economically very important parameters. They measure the impact in the current period on each endogenous variable of a unit change in any predetermined variable. Consider, for example, a unit increase in the level of g_t. From Eq. (1-8) of the structural form there would be a simultaneous increase of one unit in GNP. But from the consumption function (1-6), increases in GNP will induce increases in consumption, which in turn, from Eq. (1-8), will induce further increases in GNP. The reduced-form coefficient

$$\frac{\partial y_t}{\partial g_t} = \pi_{11} = \frac{1}{1 - \alpha_1(1 - \tau)}$$

shows the end result of this process in period t. This is the national income multiplier of simple Keynesian theory. For example, if $\tau = 0.25$ and $\alpha_1 = 0.8$, $\pi_{11} = 2.5$, so that a unit increase in government expenditure, with tax rates and all other parameters unchanged, would raise national income in the same period by 2.5 units. Similarly, an inspection of

$$\pi_{10} = (\alpha_0 + \beta_0)\delta$$

shows that a unit increase (upward shift) in the intercept of either the consumption or the investment function would have equal multiplier effects on GNP. All the π's are multipliers, and they are termed *impact multipliers*, because they show the effect in the *current* period of changes in predetermined variables. Estimates of the structural coefficients can yield estimates of the reduced-form coefficients, and so these impact multipliers can be evaluated. Alternatively, the reduced-form equations may be estimated directly. These topics will be discussed in the chapter on simultaneous equation estimation later in the book.

The impact effects in period t are not the end of the story. Let us write Eq. (1-12) in first difference form,

$$\Delta y_t = \pi_{11}\Delta g_t + \pi_{12}\Delta r_t + \pi_{13}\Delta r_{t-1} + \pi_{14}\Delta y_{t-1} + \pi_{15}\Delta y_{t-2} \qquad (1\text{-}15)$$

where

$$\Delta y_t = y_t - y_{t-1}, \ldots$$

Let us suppose that g and r have been held constant sufficiently long for y to settle down at some constant equilibrium level. This involves the implicit assumption that equilibrium values exist and that the system is stable, and we will return to this point below. Equilibrium thus implies

$$\begin{aligned} \Delta g_t &= \Delta g_{t-1} = \Delta g_{t-2} = \cdots = 0 \\ \Delta r_t &= \Delta r_{t-1} = \Delta r_{t-2} = \cdots = 0 \\ \Delta y_t &= \Delta y_{t-2} = \Delta y_{t-2} = \cdots = 0 \end{aligned}$$

Now suppose that the level of government expenditure in period $t+1$ is raised by an amount d and then held constant at the new level indefinitely, that is,

$$\Delta g_{t+1} = d, \qquad \Delta g_{t+2} = \Delta g_{t+3} = \cdots = 0$$

From Eq. (1-15) the impact effect on national income in period $t+1$ is

$$\Delta y_{t+1} = \pi_{11} d$$

In period $t+2$, Eq. (1-15) reads

$$\Delta y_{t+2} = \pi_{14}\Delta y_{t+1} = \pi_{14}\pi_{11} d$$

In period $t+3$, the equation reads

$$\begin{aligned} \Delta y_{t+3} &= \pi_{14}\Delta y_{t+2} + \pi_{15}\Delta y_{t+1} \\ &= \pi_{14}^2\pi_{11} d + \pi_{15}\pi_{11} d \end{aligned}$$

Thus the one-step change in g sets off a sequence of changes in y because of the lags in the system. There is thus a whole series of lagged multipliers, namely,

$$\frac{\partial y_{t+1}}{\partial g_{t+1}} = \pi_{11} \qquad \text{zero lag, or impact multipliers}$$

$$\frac{\partial y_{t+2}}{\partial g_{t+1}} = \pi_{14}\pi_{11} \qquad \text{one-period lag}$$

$$\frac{\partial y_{t+3}}{\partial g_{t+1}} = (\pi_{14}^2 + \pi_{15})\pi_{11} \qquad \text{two-period lag}$$

The estimated reduced form can be applied sequentially to trace out the dynamic