# A Guide to Econometrics

Peter Kennedy

ECONOMETRICS

# A Guide to Econometrics

Peter Kennedy

*Simon Fraser University*

The MIT Press

Cambridge, Massachusetts

Second printing, 1980
First MIT Press edition, 1979

*First published in England in 1979* by
Martin Robertson & Company Ltd.



Printed and bound in the United States of America

Library of Congress catalog card number: 79-63894
ISBN 0-262-11073-3 (hard)
ISBN 0-262-61031-0 (paper)

# Preface

This book is a *supplement* to econometrics texts at either the introductory or advanced level. As such its purpose is to enhance the ability of the text to communicate the subject of econometrics to its readers. Unlike most such supplements, it does not do this through new examples, applications or exercises; instead it provides an overview of the subject and an intuitive feel for its concepts and techniques, without the usual clutter of notation and technical detail that necessarily characterize an econometrics textbook.

It is often said of econometrics textbooks that their readers miss the forest for the trees. This is inevitable – the terminology and techniques that must be taught do not allow the text to convey a proper intuitive sense of 'What's it all about?' and 'How does it all fit together?' All econometrics textbooks fail to provide this overview. This is not from lack of trying – most textbooks have excellent passages containing the relevant insights and interpretations. They make good sense to instructors, but do *not* make the expected impact on the students. Why? Because these insights and interpretations are broken up, appearing throughout the book, mixed with the technical details. In their struggle to keep up with notation and to learn these technical details, students miss the overview so essential to a real understanding of those details.

This guide was reviewed in manuscript form by a large number of instructors and used by several classes at both the introductory and advanced levels. It is interesting that the majority of the instructors did not feel that this book was needed whereas the students were unanimously of the opposite opinion. To the instructors, this guide presented nothing they did not feel was already in textbooks. To the students, this guide provided an overview of the subject and non-technical descriptions of its landmarks, without distracting notation, formulae, proofs or technical details; it provided the uninitiated with a perspective from which it was possible to assimilate the detail of the textbook more easily.

A brief introductory chapter sets the stage. The second chapter discusses at some length the criteria for choosing estimators, and in doing so develops many of the basic concepts used throughout the book. The third chapter structures the basic overview of the subject matter, presenting the five assumptions of the classical linear regression model and explaining how most problems encountered in econometrics can be interpreted as a violation of one of these assumptions. The fourth chapter discusses some concepts of inference to provide a foundation for later chapters. Chapters 5 to 9 take each of the five assumptions of the classical linear regression model in turn and describe possible violations of these assumptions, implications of these violations, and means of resolving the resulting estimation problems. A final chapter examines some selected topics.

To minimize readers' distractions, there are no footnotes. All references, peripheral points and details worthy of comment are relegated to a section at the end of each chapter entitled 'General Notes'. The few technical discussions that appear in the book are placed in end-of-chapter sections entitled 'Technical Notes'. Students are advised to wait until a second or third reading of a chapter before attempting to integrate the material in the General or Technical Notes with the body of the chapter. A glossary explains common econometric terms not found in the body of this book.

No author is without debts. Thanks are due to Judy Alexander, Sandy Christensen, Larry Clark, Charles Conrod, John Cuddington, Art Goldberger, Dick Holmes, Teh Hu, Jon Nelson, Angus Oliver, Rod Peterson, Alan Sleeman and numerous students for comments both general and detailed. Expert typing was provided by Carol Eddy, Patricia Lyles and Donna Wilson. None of these people can be held responsible for any shortcomings or errors this book may contain; only the author can be blamed.

# Dedication

To ANNA and RED

who, until they discovered what an econometrician was, were very impressed that their son might become one. With apologies to K. A. C. Manderville, I draw their attention to the following, adapted from *The Undoing of Lamia Gurdleneck.*

'You haven't told me yet,' said Lady Nuttal, 'what it is your fiancé does for a living.'

'He's an econometrician,' replied Lamia, with an annoying sense of being on the defensive.

Lady Nuttal was obviously taken aback. It had not occurred to her that econometricians entered into normal social relationships. The species, she would have surmised, was perpetuated in some collateral manner, like mules.

'But Aunt Sara, it's a very interesting profession,' said Lamia warmly.

'I don't doubt it,' said her aunt, who obviously doubted it very much. 'To express anything important in mere figures is so plainly impossible that there must be endless scope for well-paid advice on how to do it. But don't you think that life with an econometrician would be rather, shall we say, humdrum?'

Lamia was silent. She felt reluctant to discuss the surprising depth of emotional possibility which she had discovered below Edward's numerical veneer.

'It's not the figures themselves,' she said finally, 'it's what you do with them that matters.'

# Contents

# 1. Introduction

## 1.1 What is Econometrics?

Strange as it may seem, there does not exist a generally accepted answer to this question. Responses vary from the silly 'Econometrics is what econometricians do' to the staid 'Econometrics is the study of the application of statistical methods to the analysis of economic phenomena', with sufficient disagreements to warrant an entire journal article devoted to this question (Tintner, 1953).

This confusion stems from the fact that econometricians wear many different hats. At times they are *mathematicians*, formulating economic theory in ways that make it appropriate for statistical testing. At times they are *accountants*, concerned with the problem of finding and collecting economic data and relating theoretical economic variables to observable ones. At times they are *applied statisticians*, spending hours with the computer trying to estimate economic relationships or predict economic events. And at times they are *theoretical statisticians*, applying their skills to the development of statistical techniques appropriate to the empirical problems characterizing the science of economics. It is to the last of these roles that the term 'econometric theory' applies, and it is on this aspect of econometrics that most textbooks on the subject focus. This guide is accordingly devoted to this 'econometric theory' dimension of econometrics, discussing the empirical problems typical of economics and the statistical techniques used to overcome these problems.

What distinguishes an econometrician from a statistician is the former's preoccupation with problems caused by violations of statisticians' standard assumptions; due to the nature of economic relationships and the lack of controlled experimentation these assumptions are seldom met. Patching up statistical methods to deal with situations frequently encountered in empirical work in economics has created a large battery of extremely sophisticated statistical techniques. In fact, econometricians are often accused of using sledgehammers to crack open peanuts while turning a blind eye to data deficiencies and the many questionable assumptions required for the successful application of these techniques. Valavanis has expressed this feeling forcefully:

> Econometric theory is like an exquisitely balanced French recipe, spelling out precisely with how many turns to mix the sauce, how many carats of spice to add, and for how many milliseconds to bake the mixture at exactly 474 degrees of temperature. But when the statistical cook turns to raw materials, he finds that hearts of cactus fruit are unavailable, so he substitutes chunks of cantaloupe; where the recipe calls for vermicelli he uses shredded wheat; and he substitutes green garment dye for curry, ping-pong balls for turtle's eggs, and, for Chalifougnac vintage 1883, a can of turpentine. [1959, p. 83]

Criticisms of econometrics along these lines are not uncommon. Rebuttals cite improvements in data collecting, extol the fruits of the computer revolution and provide examples of improvements in estimation due to advanced techniques. It remains a fact, though, that in practice good results depend as much on the input of sound and imaginative economic theory as on the application of correct statistical methods. The skill of the econometrician lies in judiciously mixing these two essential ingredients; in the words of Malinvaud:

> The art of the econometrician consists in finding the set of assumptions which are both sufficiently specific and sufficiently realistic to allow him to take the best possible advantage of the data available to him. [1966, p. 514]

Modern econometrics texts try to infuse this art into students by providing a large number of detailed examples of empirical applications. This important dimension of econometrics texts lies beyond the scope of this book. Readers should keep this in mind as they use this guide to improve their understanding of the purely statistical methods of econometrics.

## 1.2 The Disturbance Term

A major distinction between economists and econometricians is the latter's concern with disturbance terms. An economist will specify, for example, that consumption is a function of income, and write $C=f(Y)$ where $C$ is consumption and $Y$ is income. An econometrician will claim that this relationship must also include a *disturbance* (or *error*) term, and may alter the equation to read

$C = f(Y) + \varepsilon$ where $\varepsilon$ (epsilon) is a disturbance term. Without the disturbance term the relationship is said to be *exact* or *deterministic*; with the disturbance term it is said to be *stochastic*.

The word 'stochastic' comes from the Greek 'stokhos' meaning a target or bull's eye. A stochastic relationship is not always right on target in the sense that it predicts the precise value of the variable being explained, just as a dart thrown at a target seldom hits the bull's eye. The disturbance term is used to capture explicitly the size of these 'misses' or 'errors'. The existence of the disturbance term is justified in three main ways. (Note these are not mutually exclusive.)

(1) *Omission of the influence of innumerable chance events.* Although income might be the major determinant of the level of consumption, it is not the only determinant. Other variables, such as the interest rate of liquid asset holdings, may have a systematic influence on consumption. Their omission constitutes one type of *specification error*: the nature of the economic relationship is not correctly specified. In addition to these systematic influences, however, are innumerable less-systematic influences such as weather variations, taste changes, earthquakes, epidemics and postal strikes. Although some of these variables may have a significant impact on consumption, and thus should definitely be included in the specified relationship, many have only a very slight, irregular influence; the disturbance is often viewed as representing the net influence of a large number of such small and independent causes.

(2) *Measurement error.* It may be the case that the variable being explained cannot be measured accurately, either because of data collection difficulties or because it is inherently unmeasurable and a proxy variable must be used in its stead. The disturbance term can in these circumstances be thought of as representing this measurement error. But note that errors in measuring the explaining variable(s) (as opposed to the variable being explained) are treated differently under the heading *errors in variables*. The terminology *errors in equations* is sometimes used to denote errors or disturbances in the context in which they are being discussed here.

(3) *Human indeterminancy.* Some people believe that human behaviour is such that actions taken under identical circumstances will differ in a random way. The disturbance term can be thought of as representing this inherent randomness in human behaviour.

Associated with any explanatory relationship are unknown constants, called *parameters*, which tie the relevant variables into an equation. For example, the relationship between consumption and income could be specified as

$$C = \beta_1 + \beta_2 Y + \varepsilon$$

where $\beta_1$ and $\beta_2$ are the parameters characterizing this consumption function. Economists are often keenly interested in learning the values of these unknown parameters.

The existence of the disturbance term, coupled with the fact that its magnitude is unknown, makes calculation of these parameter values impossible. Instead they must be *estimated*. It is on this task, the estimation of parameter values, that the bulk of econometric theory focuses. The success of econometricians' methods of estimating parameter values depends in large part on the nature of the disturbance term; statistical assumptions concerning the characteristics of the disturbance term, and means of testing these assumptions, therefore play a prominent role in econometric theory.

### 1.3 Estimates and Estimators

In their mathematical notation, econometricians usually employ Greek letters to represent the true, unknown values of parameters. The Greek letter most often used in this context is beta ($\beta$). Thus, throughout this book, $\beta$ is used as the parameter value that the econometrician is seeking to learn. Of course, no one ever actually learns the value of $\beta$, but it can be estimated: via statistical techniques empirical data can be used to take an educated guess at $\beta$. In any particular application, an estimate of $\beta$ is simply a number. For example, $\beta$ might be estimated as 16.2. But in general econometricians are seldom interested in estimating a single parameter; economic relationships are usually sufficiently complex as to require more than one parameter, and because these parameters occur in the same relationship better estimates of these parameters can be obtained if they are estimated together (i.e., the influence of one explaining variable is more accurately captured if the influence of the other explaining variables is simultaneously accounted for). As a result, $\beta$ seldom refers to a single parameter value; it almost always refers to a set of parameter values, individually called $\beta_1, \beta_2, \ldots, \beta_k$ where $k$ is the number of different parameters in the set. $\beta$ is then referred to as a vector and is written as

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$$

In any particular application, an estimate of β will be a set of numbers. For example, if three parameters are being estimated (i.e., the dimension of β is three), β might be estimated as

$$\begin{bmatrix} 0.8 \\ 1.2 \\ -4.6 \end{bmatrix}$$

In general, econometric theory focuses not on the estimate itself, but on the *estimator* – the formula or 'recipe' by which the data are transformed into an actual estimate. The reason for this is that the justification of an estimate computed from a particular sample rests on a justification of the estimation method (the estimator). The econometrician has no way of knowing the actual values of the disturbances inherent in a sample of data; depending on these disturbances, an estimate calculated from that sample could be quite inaccurate. It is therefore impossible to justify the estimate itself. However, it may be the case that the econometrician can justify the estimator by showing, for example, that that estimator 'usually' produces an estimate that is 'quite close' to the true parameter value regardless of the particular sample chosen. (The meaning of this sentence is discussed at length in the next chapter.) Thus an estimate of β from a particular sample is defended by justifying the estimator.

Because attention is focused on estimators of β, a convenient way of denoting those estimators is required. An easy way of doing this is to place a mark over the β or a superscript on it. Thus $\hat{\beta}$ (beta-hat) and $\beta^*$ (beta-star) are often used to denote estimators of beta. One estimator, the ordinary least squares (OLS) estimator, is very popular in econometrics; the notation $\beta^{OLS}$ is used throughout this book to represent it. Alternative estimators are denoted by $\hat{\beta}$, $\beta^*$, or something similar.

## 1.4 Good and Preferred Estimators

Any fool can produce an estimator of β, since literally an infinite number of them exists, i.e., there exists an infinite number of different ways in which a sample of data can be used to produce an estimate of β, all but a few of these ways producing 'bad' estimates. What distinguishes an econometrician is the ability to produce 'good' estimators, which in turn produce 'good' estimates. One of these 'good' estimators could be chosen as the 'best' or 'preferred' estimator and be used to generate the 'preferred' estimate of β. What further distinguishes an econometrician is his ability to provide 'good' estimators in a variety of different estimating contexts. The set of 'good' estimators (and the choice of 'preferred' estimator) is not the same in all estimating problems. In fact, a 'good' estimator in one estimating situation could be a 'bad' estimator in another situation.

The study of econometrics revolves around how to generate a 'good' or the 'preferred' estimator in a given estimating situation. But before the 'how to' can be explained, the meaning of 'good' and 'preferred' must be made clear. This takes the discussion into the subjective realm: the meaning of 'good' or 'preferred' estimator depends upon the subjective values of the person doing the estimating. The best the econometrician can do under these circumstances is to recognize the more popular criteria used in this regard and generate estimators that meet one or more of these criteria. Estimators meeting certain of these criteria could be called 'good' estimators. The ultimate choice of the 'preferred' estimator, however, lies in the hands of the person doing the estimating, for it is his or her value judgments that determine which of these criteria is the most important. This value judgment may well be influenced by the purpose for which the estimate is sought, in addition to the subjective prejudices of the individual.

Clearly, our investigation of the subject of econometrics can go no further until the possible criteria for a 'good' estimator are discussed. This is the purpose of the next chapter.

# General Notes

### 1.1

- The term 'econometrics' first came into prominence with the formation in the early 1930s of the Econometric Society and the founding of the journal *Econometrica*. The introduction of Dowling and Glahe (1970) surveys briefly the landmark publications related to econometrics.
- Brunner (1973), Rubner (1970) and Streissler (1970) are good sources of cynical views of econometrics. More comments appear in this book in section 8A.2 on errors in variables and section 10.5 on prediction. Fair

(1973) and Fromm and Schink (1973) are examples of studies defending the use of sophisticated econometric techniques.

- Critics might choose to paraphrase the Malinvaud quote as 'The art of drawing a crooked line from an unproved assumption to a foregone conclusion'. The importance of a proper understanding of econometric techniques in the face of a potential inferiority of econometrics to inspired economic theorizing is captured nicely by Samuelson (1965) p. 9: 'Even if a scientific regularity were less accurate than the intuitive hunches of a virtuoso, the fact that it can be put into operation by thousands of people who are not virtuosos gives it a transcendental importance.' This guide is designed for those of us who are not virtuosos!
- There exist several books addressed to the empirical applications dimension of econometrics. Examples are Bridge (1971), Cramer (1971), Desai (1976), Pindyck and Rubinfeld (1976), Wallis (1973) and Wynn and Holden (1974). Evans (1969) is also useful in this regard. In addition, the econometric theory texts of Gujarati (1978), Intriligator (1978), Koutsoyiannis (1977) and Maddala (1977) all contain numerous examples of empirical applications. Zellner (1968) contains many well-known articles addressed to this empirical dimension of econometrics.

**1.2**

- The error term associated with a relationship need not necessarily be additive, as it is in the example cited. For some non-linear functions it is often convenient to specify the error term in a multiplicative form, for example, as noted later in section 5.3 on non-linear functional forms. In other instances it may be appropriate to build the stochastic element into the relationship by specifying the parameters to be random variables rather than constants. (This is called the random-coefficients model.) These are usually treated as special topics in econometrics.
- Econometricians usually do not know the actual form of the economic relationship being studied and consequently commit a specification error when formulating the estimation problem, either by omitting variables that should be included or by adopting an incorrect functional form. In these cases the econometrician often views the disturbance term as incorporating this specification error in addition to the other types of error cited. Following this operational procedure creates an error term with unusual properties; estimation under these circumstances is discussed under the heading of specification error (chapter 5).
- Estimation of parameter values is not the only purpose of econometrics. Two other major themes can be identified: testing of hypotheses and economic forecasting. Because both these problems are intimately related to the estimation of parameter values, it is not misleading to characterize econometrics as being primarily concerned with parameter estimation.
- In terms of the throwing-darts-at-a-target analogy, characterizing disturbance terms refers to describing the nature of the misses: are the darts distributed uniformly around the bull's eye? Is the average miss large or small? Does the average miss depend on who is throwing the darts? Is a

miss to the right likely to be followed by another miss to the right? In later chapters the statistical specification of these characteristics and the related terminology (such as 'homoskedasticity' and 'autocorrelated errors') are explained in considerable detail.

## 1.3

- An estimator is simply an algebraic function of a potential sample of data; once the sample is drawn, this function creates an actual numerical estimate.
- Chapter 2 discusses in detail the means whereby an estimator is 'justified' and compared to alternative estimators.

## 1.4

- The terminology 'preferred' estimator is used instead of the term 'best' estimator because the latter has a specific meaning in econometrics. This is explained in chapter 2.