A GUIDE TO ECONOMETRICS

FIFTH EDITION

PETER KENNEDY

Simon Fraser University

Blackwell Publishing

BLACKWELL PUBLISHING
350 Main Street, Malden, MA 02148-5020, USA
9600 Garsington Road, Oxford OX4 2DQ, UK
550 Swanston Street, Carlton, Victoria 3053, Australia

First published 1979 by Martin Robertson and Company Ltd
Second and third editions published by Blackwell Publishers Ltd
Fourth edition published 1998
Fifth edition published 2003 by Blackwell Publishing Ltd

4 2006

Library of Congress Cataloging-in-Publication Data

Kennedy, Peter, 1943–
A guide to econometrics / Peter Kennedy. – 5th ed.
p. cm.
ISBN 1-4051-1501-7 (hardcover : alk. paper) – ISBN 1-4051-1502-5 (pbk. : alk. paper)
1. Econometrics. I. Title.

HB139.K45 2003
330'.01'5195–dc21
2002156370

ISBN-13: 978-1-4051-1501-8 (hardcover : alk. paper) – ISBN-13: 978-1-4051-1502-5 (pbk. : alk. paper)

A catalogue record for this title is available from the British Library.

Set in 10/12pt Baskerville
by Graphicraft Ltd, Hong Kong
Printed and bound in India
by Replika Press Pvt Ltd, Kundli

The publisher's policy is to use permanent paper from mills that operate a sustainable forestry policy, and which has been manufactured from pulp processed using acid-free and elementary chlorine-free practices. Furthermore, the publisher ensures that the text paper and cover board used have met acceptable environmental accreditation standards.

For further information on

CONTENTS

PREFACE

Upper-level undergraduate and beginning graduate econometrics students have found the previous editions of this book to be of immense value to their understanding of econometrics. And judging by sales, more and more instructors of econometrics have come to recognize this, so that students are as likely to learn about this book from a course outline as from word of mouth, the phenomenon that made the early editions of this book so successful.

What is it about this book that students have found to be of such value? This book supplements econometrics texts, at all levels, by providing 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 they 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 book provides students with a perspective from which it is possible to assimilate more easily the details of these textbooks.

Although the changes from the fourth edition are numerous, the basic structure and flavor of the book remain unchanged. Following an introductory chapter, 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 provides an 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 exposits some concepts of inference to provide a foundation for later chapters. Chapter 5 discusses general approaches to the specification of an econometric model, setting the stage for the next six chapters, each of which deals with violations of an assumption of the classical linear regression model, describes their implications, discusses relevant tests, and suggests means of resolving resulting estimation problems. The remaining ten chapters and appendices A, B, and C address selected topics. Appendix D provides some student exercises and Appendix E offers suggested answers to the even-numbered exercises. A glossary explains common econometric terms not found in the body of the book. A set of suggested answers to odd-numbered questions is available from the publisher upon request to instructors adopting this book for classroom use.

This edition is a major revision, primarily because of two new chapters, on panel data (chapter 17) and on applied econometrics (chapter 21). Earlier editions were missing the former because of my laziness, and the latter because it was too hard to write. Both retain the flavor of the book, with the latter being quite unique in this regard. (Econometrics texts do not exposit the ten commandments of applied econometrics, discuss reasons for getting the "wrong" sign, or list common mistakes made by practitioners, for example.) Several other chapters have had extensive revision, most notably the specification chapter to fit with the new applied econometrics chapter, the qualitative dependent variables chapter to improve the discussion of multinomial/conditional logit, the limited dependent variables chapter to exposit more carefully the context for application of the Tobit model, and the time series chapter to incorporate a section on vector autoregression (transplanted from the simultaneous equations chapter) and to describe more fully the structure of the vector error correction model. In the exercises, two new sections of questions have been added, on applied econometrics and on bootstrapping. Innumerable additions and changes, major and minor, have been made throughout to update results and references, and to improve exposition.

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 technical material that appears in the book is placed in end-of-chapter sections entitled "Technical Notes." This technical material continues to be presented in a way that supplements rather than duplicates the contents of traditional textbooks. Students should find that this material provides a useful introductory bridge to the more sophisticated presentations found in their main text.

Errors in or shortcomings of this book are my responsibility, but for improvements I owe many debts, mainly to scores of students, both graduate and undergraduate, whose comments and reactions have played a prominent role in shaping this fifth edition. More than a dozen anonymous referees reviewed this edition, many of them providing detailed suggestions for improvement, some of which, but not all, were incorporated. I continue to be grateful to students throughout the world who have expressed thanks to me for writing this book; I hope this fifth edition continues to be of value to students both during and after their formal course-work.

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. First, and foremost, they are *economists*, capable of utilizing economic theory to improve their empirical analyses of the problems they address. 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; owing 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 die for curry, ping-pong balls for turtle's eggs, and, for Chalifougnac vintage 1883, a can of turpentine. (Valavanis, 1959, p. 83)

How has this state of affairs come about? One reason is that prestige in the econometrics profession hinges on technical expertise rather than on the hard work required to collect good data:

> It is the preparation skill of the econometric chef that catches the professional eye, not the quality of the raw materials in the meal, or the effort that went into procuring them. (Griliches, 1994, p. 14)

Criticisms of econometrics along these lines are not uncommon. Rebuttals cite improvements in data collection, 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. (Malinvaud, 1966, p. 514)

Modern econometrics texts try to infuse this art into students by providing a large number of detailed examples of empirical application. 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 ε (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 or 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. Errors in measuring the explaining variable(s) (as opposed to the variable being explained) create a serious econometric problem, discussed in chapter 9. The terminology *errors in variables* is also used to refer to measurement errors.

(3) *Human indeterminacy* Some people believe that human behavior 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 behavior.

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 β_1 and β_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 (β). Thus, throughout this book, β is used as the parameter value that the econometrician is seeking to learn. Of course, no one ever actually learns the value of β, but it can be estimated: via statistical techniques, empirical data can be used to take an educated guess at β. In any particular application, an estimate of β is simply a number. For example, β might be estimated as 16.2. But, in general, econometricians are seldom interested in estimating a single parameter; economic relationships are usually sufficiently complex 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, β 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. β 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., if 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 the 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, in particular the meaning of "usually" and of "quite close," 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 β^* (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 β^{OLS} is used throughout this book to represent it. Alternative estimators are denoted by $\hat{\beta}$, β^*, or something similar. Many textbooks use the letter b to denote the OLS estimator.

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 the 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.

2 CRITERIA FOR ESTIMATORS

2.1 INTRODUCTION

Chapter 1 posed the question, What is a "good" estimator? The aim of this chapter is to answer that question by describing a number of criteria that econometricians feel are measures of "goodness." These criteria are discussed under the following headings:

(1) Computational cost
(2) Least squares
(3) Highest R^2
(4) Unbiasedness
(5) Efficiency
(6) Mean square error
(7) Asymptotic properties
(8) Maximum likelihood

Discussion of one major criterion, robustness (insensitivity to violations of the assumptions under which the estimator has desirable properties as measured by the criteria above), is postponed to chapter 20. Since econometrics can be characterized as a search for estimators satisfying one or more of these criteria, care is taken in the discussion of the criteria to ensure that the reader understands fully the meaning of the different criteria and the terminology associated with them. Many fundamental ideas of econometrics, critical to the question, "What's econometrics all about?", are presented in this chapter.

2.2 COMPUTATIONAL COST

To anyone, but particularly to economists, the extra benefit associated with choosing one estimator over another must be compared with its extra cost, where cost refers to expenditure of both money and effort. Thus, the computational ease and cost of using one estimator rather than another must be taken into account whenever selecting an estimator. Fortunately, the existence and ready availability of high-speed computers, along with standard packaged routines for most of the popular estimators, has made computational cost very low. As a result, this criterion does not play as strong a role as it once did. Its influence is now felt only when dealing with two kinds of estimators. One is the case of an atypical estimation procedure for which there does not exist a readily available packaged computer program and for which the cost of programming is high. The second is an estimation method for which the cost of running a packaged program is high because it needs large quantities of computer time; this could occur, for example, when using an iterative routine to find parameter estimates for a problem involving several nonlinearities.

2.3 LEAST SQUARES

For any set of values of the parameters characterizing a relationship, estimated values of the dependent variable (the variable being explained) can be calculated using the values of the independent variables (the explaining variables) in the data set. These estimated values (called $\hat{y}$) of the dependent variable can be subtracted from the actual values (y) of the dependent variable in the data set to produce what are called the *residuals* ($y - \hat{y}$). These residuals could be thought of as estimates of the unknown disturbances inherent in the data set. This is illustrated in figure 2.1. The line labeled $\hat{y}$ is the estimated relationship corresponding to a specific set of values of the unknown parameters. The dots represent actual observations on the dependent variable y and the independent variable x. Each observation is a certain vertical distance away from the estimated line, as pictured by the double-ended arrows. The lengths of these double-ended arrows measure the residuals. A different set of specific values of the parameters would create a different estimating line and thus a different set of residuals.

It seems natural to ask that a "good" estimator be one that generates a set of estimates of the parameters that makes these residuals "small." Controversy arises, however, over the appropriate definition of "small." Although it is agreed that the estimator should be chosen to minimize a weighted sum of all these residuals, full agreement as to what the weights should be does not exist. For example, those feeling that all residuals should be weighted equally advocate choosing the estimator that minimizes the sum of the absolute values of these residuals. Those feeling that large residuals should be avoided advocate weighting larger residuals more

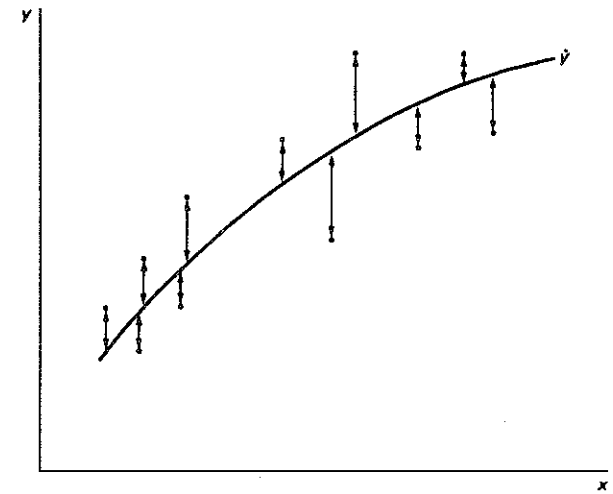

Figure 2.1 Minimizing the sum of squared residuals

heavily by choosing the estimator that minimizes the sum of the squared values of these residuals. Those worried about misplaced decimals and other data errors advocate placing a constant (sometimes zero) weight on the squared values of particularly large residuals. Those concerned only with whether or not a residual is bigger than some specified value suggest placing a zero weight on residuals smaller than this critical value and a weight equal to the inverse of the residual on residuals larger than this value. Clearly a large number of alternative definitions could be proposed, each with appealing features.

By far the most popular of these definitions of "small" is the minimization of the sum of squared residuals. The estimator generating the set of values of the parameters that minimizes the sum of squared residuals is called the *ordinary least squares* estimator. It is referred to as the OLS estimator and is denoted by β^{OLS} in this book. This estimator is probably the most popular estimator among researchers doing empirical work. The reason for this popularity, however, does *not* stem from the fact that it makes the residuals "small" by minimizing the sum of squared residuals. Many econometricians are leery of this criterion because minimizing the sum of squared residuals does not say anything specific about the relationship of the estimator to the true parameter value β that it is estimating. In fact, it is possible to be too successful in minimizing the sum of squared residuals, accounting for so many unique features of that *particular sample* that the estimator loses its general validity, in the sense that, were that estimator applied to a new sample, poor estimates would result. The great popularity of the OLS estimator comes from the fact that in some estimating problems (but not all!) it scores well on some of the other criteria, described below, that are thought to be of greater importance. A secondary reason for its popularity is its computational ease; all computer packages include the OLS estimator for linear relationships, and many have routines for nonlinear cases.

Because the OLS estimator is used so much in econometrics, the characteristics of this estimator in different estimating problems are explored very thoroughly by all econometrics texts. The OLS estimator *always* minimizes the sum of squared residuals; but it does *not* always meet other criteria that econometricians feel are more important. As will become clear in the next chapter, the subject of econometrics can be characterized as an attempt to find alternative estimators to the OLS estimator for situations in which the OLS estimator does not meet the estimating criterion considered to be of greatest importance in the problem at hand.

2.4 HIGHEST R^2

A statistic that appears frequently in econometrics is the coefficient of determination, R^2. It is supposed to represent the proportion of the variation in the dependent variable "explained" by variation in the independent variables. It does this in a meaningful sense in the case of a linear relationship estimated by OLS. In this case it happens that the sum of the squared deviations of the dependent variable about its mean (the "total" variation in the dependent variable) can be broken into two parts, called the "explained" variation (the sum of squared deviations of the estimated values of the dependent variable around their mean) and the "unexplained" variation (the sum of squared residuals). R^2 is measured either as the ratio of the "explained" variation to the "total" variation or, equivalently, as 1 minus the ratio of the "unexplained" variation to the "total" variation, and thus represents the percentage of variation in the dependent variable "explained" by variation in the independent variables.

Because the OLS estimator minimizes the sum of squared residuals (the "unexplained" variation), it automatically maximizes R^2. Thus maximization of R^2, as a criterion for an estimator, is formally identical to the least squares criterion, and as such it really does not deserve a separate section in this chapter. It is given a separate section for two reasons. The first is that the formal identity between the highest R^2 criterion and the least squares criterion is worthy of emphasis. And the second is to distinguish clearly the difference between applying R^2 as a criterion in the context of searching for a "good" estimator when the functional form and included

independent variables are known, as is the case in the present discussion, and using R^2 to help determine the proper functional form and the appropriate independent variables to be included. This latter use of R^2, and its misuse, are discussed later in the book (in sections 5.5 and 6.2).

2.5 UNBIASEDNESS

Suppose we perform the conceptual experiment of taking what is called a *repeated* sample: keeping the values of the independent variables unchanged, we obtain new observations for the dependent variable by drawing a new set of disturbances. This could be repeated, say, 2,000 times, obtaining 2,000 of these repeated samples. For each of these repeated samples we could use an estimator β^* to calculate an estimate of β. Because the samples differ, these 2,000 estimates will not be the same. The manner in which these estimates are distributed is called the *sampling distribution* of β^*. This is illustrated for the one-dimensional case in figure 2.2, where the sampling distribution of the estimator is labeled $f(\beta^*)$. It is simply the probability density function of β^*, approximated by using the 2,000 estimates of β to construct a histogram, which in turn is used to approximate the relative frequencies of different estimates of β from the estimator β^*. The sampling distribution of an alternative estimator, $\hat{\beta}$, is also shown in figure 2.2.

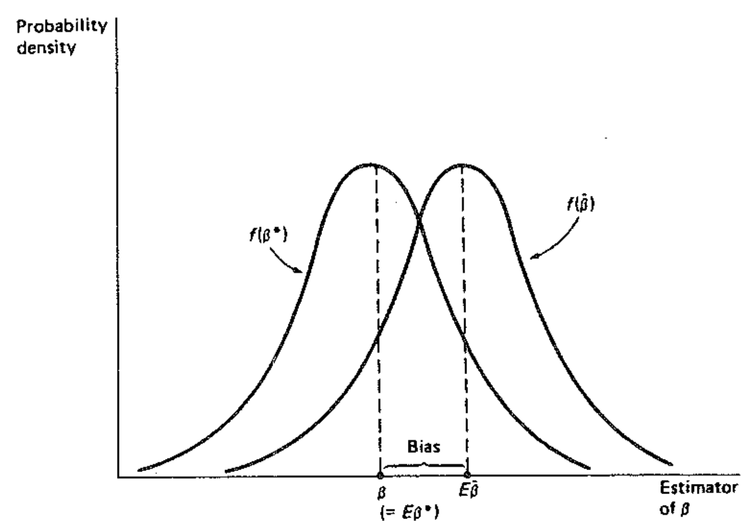

Figure 2.2 Using the sampling distribution to illustrate bias

This concept of a sampling distribution, the distribution of estimates produced by an estimator in repeated sampling, is crucial to an understanding of econometrics. Appendix A at the end of this book discusses sampling distributions at greater length. Most estimators are adopted because their sampling distributions have "good" properties; the criteria discussed in this and the following three sections are directly concerned with the nature of an estimator's sampling distribution.

The first of these properties is unbiasedness. An estimator β^* is said to be an *unbiased* estimator of β if the mean of its sampling distribution is equal to β, i.e., if the average value of β^* in repeated sampling is β. The mean of the sampling distribution of β^* is called the expected value of β^* and is written $E\beta^*$; the bias of β^* is the difference between $E\beta^*$ and β. In figure 2.2, β^* is seen to be unbiased, whereas $\hat{\beta}$ has a bias of size $(E\hat{\beta} - \beta)$. The property of unbiasedness does not mean that $\beta^* = \beta$; it says only that, if we could undertake repeated sampling an infinite number of times, we would get the correct estimate "on the average."

The OLS criterion can be applied with no information concerning how the data were generated. This is not the case for the unbiasedness criterion (and all other criteria related to the sampling distribution), since this knowledge is required to construct the sampling distribution. Econometricians have therefore developed a standard set of assumptions (discussed in chapter 3) concerning the way in which observations are generated. The general, but not the specific, way in which the disturbances are distributed is an important component of this. These assumptions are sufficient to allow the basic nature of the sampling distribution of many estimators to be calculated, either by mathematical means (part of the technical skill of an econometrician) or, failing that, by an empirical means called a Monte Carlo study, discussed in section 2.10.

Although the mean of a distribution is not necessarily the ideal measure of its location (the median or mode in some circumstances might be considered superior), most econometricians consider unbiasedness a desirable property for an estimator to have. This preference for an unbiased estimator stems from the *hope* that a particular estimate (i.e., from the sample at hand) will be close to the mean of the estimator's sampling distribution. Having to justify a particular estimate on a "hope" is not especially satisfactory, however. As a result, econometricians have recognized that being centered over the parameter to be estimated is only *one* good property that the sampling distribution of an estimator can have. The variance of the sampling distribution, discussed next, is also of great importance.

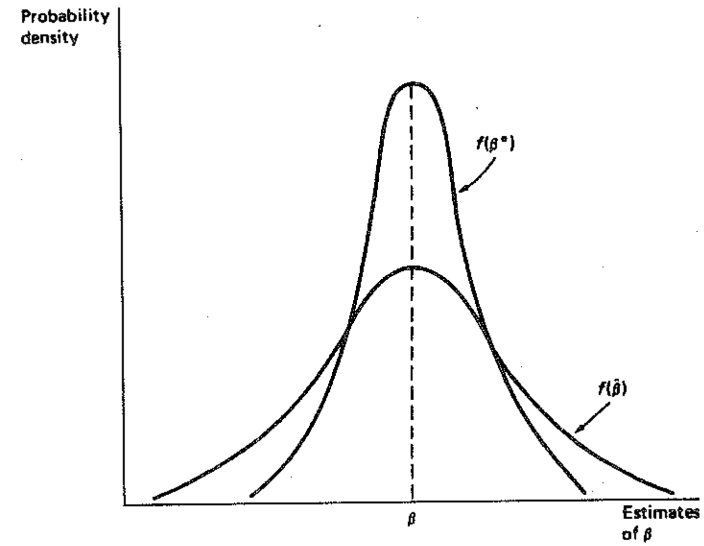

Figure 2.3 Using the sampling distribution to illustrate efficiency

2.6 EFFICIENCY

In some econometric problems it is impossible to find an unbiased estimator. But whenever one unbiased estimator can be found, it is usually the case that a large number of other unbiased estimators can also be found. In this circumstance the unbiased estimator whose sampling distribution has the smallest variance is considered the most desirable of these unbiased estimators; it is called the *best unbiased* estimator, or the *efficient* estimator among all unbiased estimators. Why it is considered the most desirable of all unbiased estimators is easy to visualize. In figure 2.3 the sampling distributions of two unbiased estimators are drawn. The sampling distribution of the estimator $\hat{\beta}$, denoted $f(\hat{\beta})$, is drawn "flatter" or "wider" than the sampling distribution of β^*, reflecting the larger variance of $\hat{\beta}$. Although both estimators would produce estimates in repeated samples whose average would be β, the estimates from $\hat{\beta}$ would range more widely and thus would be less desirable. A researcher using $\hat{\beta}$ would be less certain that his or her estimate was close to β than would a researcher using β^*.

Sometimes reference is made to a criterion called "minimum variance." This criterion, by itself, is meaningless. Consider the estimator $\beta^* = 5.2$ (i.e., whenever a sample is taken, estimate β by 5.2 ignoring the sample). This estimator has a variance of zero, the smallest possible variance, but no one would use this estimator because it performs so poorly on other criteria such as unbiasedness. (It is interesting to note, however, that it performs exceptionally well on the computational cost criterion!) Thus, whenever the minimum variance, or "efficiency," criterion is mentioned, there must exist, at least implicitly, some additional constraint, such as unbiasedness, accompanying that criterion. When the additional constraint accompanying the minimum variance criterion is that the estimators under consideration be unbiased, the estimator is referred to as the *best unbiased* estimator.

Unfortunately, in many cases it is impossible to determine mathematically which estimator, of all unbiased estimators, has the smallest variance. Because of this problem, econometricians frequently add the further restriction that the estimator be a *linear* function of the observations on the dependent variable. This reduces the task of finding the efficient estimator to mathematically manageable proportions. An estimator that is linear and unbiased and that has minimum variance among all linear unbiased estimators is called the *best linear unbiased estimator* (BLUE). The BLUE is very popular among econometricians.

This discussion of minimum variance or efficiency has been implicitly undertaken in the context of a unidimensional estimator, i.e., the case in which β is a single number rather than a vector containing several numbers. In the multidimensional case the variance of $\hat{\beta}$ becomes a matrix called the variance–covariance matrix of $\hat{\beta}$. This creates special problems in determining which estimator has the smallest variance. The technical notes to this section discuss this further.

2.7 MEAN SQUARE ERROR (MSE)

Using the best unbiased criterion allows unbiasedness to play an extremely strong role in determining the choice of an estimator, since only unbiased estimators are considered. It may well be the case that, by restricting attention to only unbiased estimators, we are ignoring estimators that are only slightly biased but have considerably lower variances. This phenomenon is illustrated in figure 2.4. The sampling distribution of $\hat{\beta}$, the best unbiased estimator, is labeled $f(\hat{\beta})$. β^* is a biased estimator with sampling distribution $f(\beta^*)$. It is apparent from figure 2.4 that, although $f(\beta^*)$ is not centered over β, reflecting the bias of β^*, it is "narrower" than $f(\hat{\beta})$, indicating a smaller variance. It should be clear from the diagram that most researchers would probably choose the biased estimator β^* in preference to the best unbiased estimator $\hat{\beta}$.

This trade-off between low bias and low variance is formalized by using as a criterion the minimization of a weighted average of the bias and the variance (i.e., choosing the estimator that minimizes this weighted average). This is not a viable formalization, however, because the bias could be

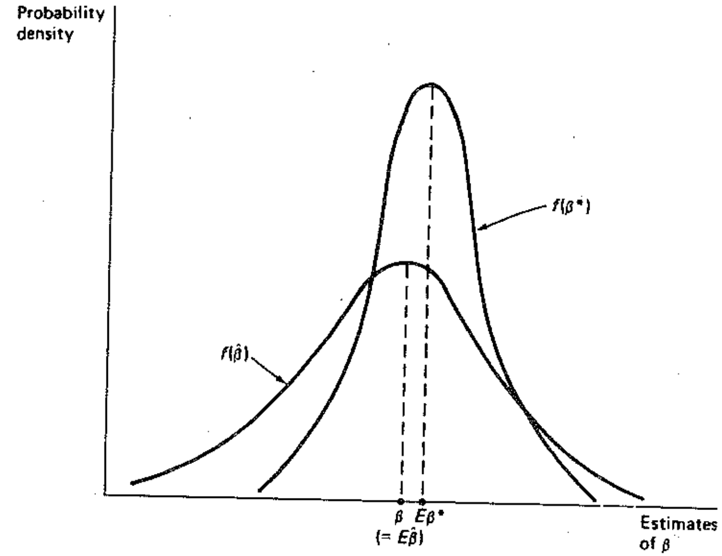

Figure 2.4 MSE trades off bias and variance

negative. One way to correct for this is to use the absolute value of the bias; a more popular way is to use its square. When the estimator is chosen so as to minimize a weighted average of the variance and the square of the bias, the estimator is said to be chosen on the *weighted square error* criterion. When the weights are equal, the criterion is the popular mean square error (MSE) criterion. The popularity of the mean square error criterion comes from an alternative derivation of this criterion: it happens that the expected value of a loss function consisting of the square of the difference between β and its estimate (i.e., the square of the estimation error) is the same as the sum of the variance and the squared bias. Minimization of the expected value of this loss function makes good intuitive sense as a criterion for choosing an estimator.

In practice, the MSE criterion is not usually adopted unless the best unbiased criterion is unable to produce estimates with small variances. The problem of multicollinearity, discussed in chapter 11, is an example of such a situation.

2.8 ASYMPTOTIC PROPERTIES

The estimator properties discussed in sections 2.5, 2.6, and 2.7 above relate to the nature of an estimator's sampling distribution. An unbiased estimator, for example, is one whose sampling distribution is centered over the true value of the parameter being estimated. These properties do not depend on the size of the sample of data at hand: an unbiased estimator, for example, is unbiased in both small and large samples. In many econometric problems, however, it is impossible to find estimators possessing these desirable sampling distribution properties in small samples. When this happens, as it frequently does, econometricians may justify an estimator on the basis of its *asymptotic* properties – the nature of the estimator's sampling distribution in extremely large samples.

The sampling distribution of most estimators changes as the sample size changes. The sample mean statistic, for example, has a sampling distribution that is centered over the population mean but whose variance becomes smaller as the sample size becomes larger. In many cases it happens that a biased estimator becomes less and less biased as the sample size becomes larger and larger – as the sample size becomes larger its sampling distribution changes, such that the mean of its sampling distribution shifts closer to the true value of the parameter being estimated. Econometricians have formalized their study of these phenomena by structuring the concept of an *asymptotic distribution* and defining desirable asymptotic or "large-sample properties" of an estimator in terms of the character of its asymptotic distribution. The discussion below of this concept and how it is used is heuristic (and not technically correct); a more formal exposition appears in appendix C at the end of this book.

Consider the sequence of sampling distributions of an estimator $\hat{\beta}$, formed by calculating the sampling distribution of $\hat{\beta}$ for successively larger sample sizes. If the distributions in this sequence become more and more similar in form to some specific distribution (such as a normal distribution) as the sample size becomes extremely large, this specific distribution is called the asymptotic distribution of $\hat{\beta}$. Two basic estimator properties are defined in terms of the asymptotic distribution.

(1) If the asymptotic distribution of $\hat{\beta}$ becomes concentrated on a particular value k as the sample size approaches infinity, k is said to be the *probability limit* of $\hat{\beta}$ and is written plim $\hat{\beta} = k$; if plim $\hat{\beta} = \beta$, then $\hat{\beta}$ is said to be *consistent*.

(2) The variance of the asymptotic distribution of $\hat{\beta}$ is called the *asymptotic variance* of $\hat{\beta}$; if $\hat{\beta}$ is consistent and its asymptotic variance is smaller than the asymptotic variance of all other consistent estimators, $\hat{\beta}$ is said to be *asymptotically efficient*.

At considerable risk of oversimplification, the plim can be thought of as the large-sample equivalent of the expected value, and so plim $\hat{\beta} = \beta$ is the large-sample equivalent of unbiasedness. Consistency can be crudely

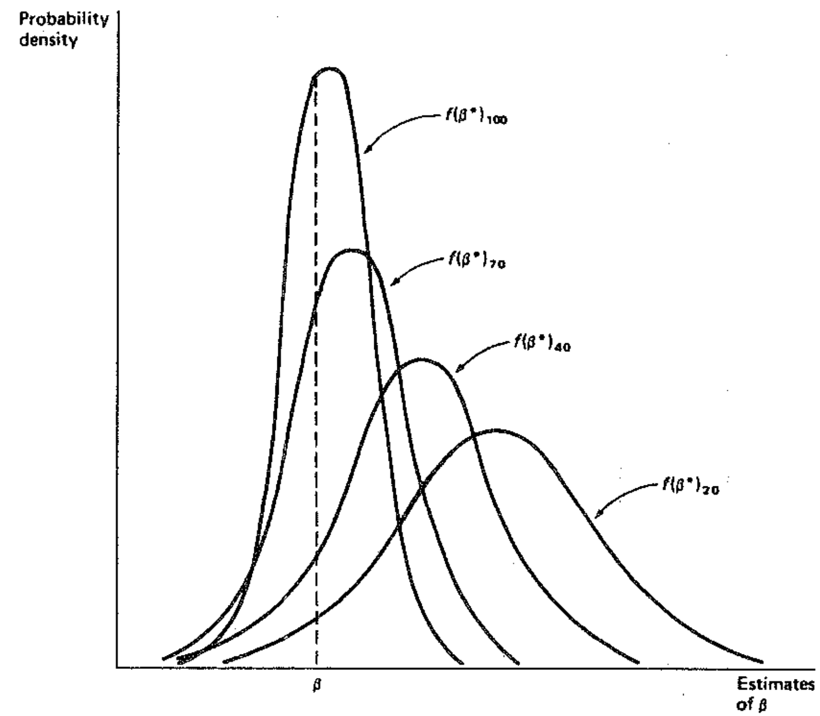

Figure 2.5 How sampling distribution can change as the sample size grows

conceptualized as the large-sample equivalent of the minimum mean square error property, since a consistent estimator can be (loosely speaking) thought of as having, in the limit, zero bias and a zero variance. Asymptotic efficiency is the large-sample equivalent of best unbiasedness: the variance of an asymptotically efficient estimator goes to zero faster than the variance of any other consistent estimator.

Figure 2.5 illustrates the basic appeal of asymptotic properties. For sample size 20, the sampling distribution of β^* is shown as $f(\beta^*)_{20}$. Since this sampling distribution is not centered over β, the estimator β^* is biased. As shown in figure 2.5, however, as the sample size increases to 40, then 70 and then 100, the sampling distribution of β^* shifts so as to be more closely centered over β (i.e., it becomes less biased), and it becomes less spread out (i.e., its variance becomes smaller). If β^* were consistent, as the sample size increased to infinity the sampling distribution would shrink in width to a single vertical line, of infinite height, placed exactly at the point β.

It must be emphasized that these asymptotic criteria are only employed in situations in which estimators with the traditional desirable small-sample properties, such as unbiasedness, best unbiasedness, and minimum mean square error, cannot be found. Since econometricians quite often must work with small samples, defending estimators on the basis of their asymptotic properties is legitimate only if it is the case that estimators with desirable asymptotic properties have more desirable small-sample properties than do estimators without desirable asymptotic properties. Monte Carlo studies (see section 2.10) have shown that in general this supposition is warranted.

The message of the discussion above is that when estimators with attractive small-sample properties cannot be found one may wish to choose an estimator on the basis of its large-sample properties. There is an additional reason for interest in asymptotic properties, however, of equal importance. Often the derivation of small-sample properties of an estimator is algebraically intractable, whereas derivation of large-sample properties is not. This is because, as explained in the technical notes, the expected value of a nonlinear function of a statistic is not the nonlinear function of the expected value of that statistic, whereas the plim of a nonlinear function of a statistic is equal to the nonlinear function of the plim of that statistic.

These two features of asymptotics give rise to the following four reasons for why asymptotic theory has come to play such a prominent role in econometrics.

(1) When no estimator with desirable small-sample properties can be found, as is often the case, econometricians are forced to choose estimators on the basis of their asymptotic properties. An example is the choice of the OLS estimator when a lagged value of the dependent variable serves as a regressor. See chapter 9.

(2) Small-sample properties of some estimators are extraordinarily difficult to calculate, in which case using asymptotic algebra can provide an indication of what the small-sample properties of this estimator are likely to be. An example is the plim of the OLS estimator in the simultaneous equations context. See chapter 10.

(3) Formulas based on asymptotic derivations are useful approximations to formulas that otherwise would be very difficult to derive and estimate. An example is the formula in the technical notes used to estimate the variance of a nonlinear function of an estimator.

(4) Many useful estimators and test statistics might never have been found had it not been for algebraic simplifications made possible by asymptotic algebra. An example is the development of LR, W, and LM test statistics for testing nonlinear restrictions. See chapter 4.

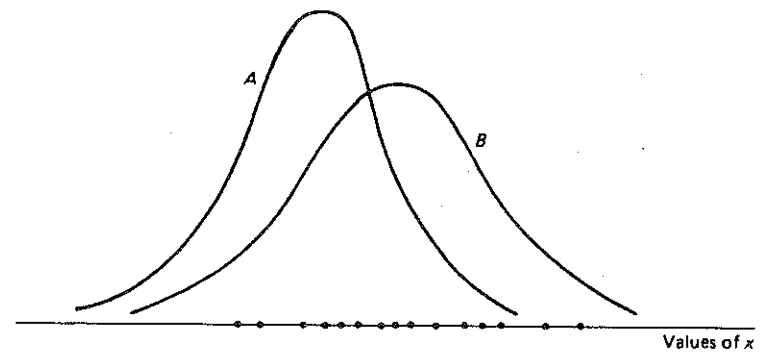

Figure 2.6 Maximum likelihood estimation

2.9 MAXIMUM LIKELIHOOD

The maximum likelihood principle of estimation is based on the idea that the sample of data at hand is more likely to have come from a "real world" characterized by one particular set of parameter values than from a "real world" characterized by any other set of parameter values. The maximum likelihood estimate (MLE) of a vector of parameter values β is simply the particular vector β^{MLE} that gives the greatest probability of obtaining the observed data.

This idea is illustrated in figure 2.6. Each of the dots represents an observation on x drawn at random from a population with mean μ and variance σ^2. Pair A of parameter values, μ^A and $(\sigma^2)^A$, gives rise in figure 2.6 to the probability density function A for x, while the pair B, μ^B and $(\sigma^2)^B$, gives rise to probability density function B. Inspection of the diagram should reveal that the probability of having obtained the sample in question if the parameter values were μ^A and $(\sigma^2)^A$ is very low compared with the probability of having obtained the sample if the parameter values were μ^B and $(\sigma^2)^B$. On the maximum likelihood principle, pair B is preferred to pair A as an estimate of μ and σ^2. The maximum likelihood estimate is the particular pair of values μ^{MLE} and $(\sigma^2)^{MLE}$ that creates the greatest probability of having obtained the sample in question; i.e., no other pair of values would be preferred to this maximum likelihood pair, in the sense that pair B is preferred to pair A. The means by which the econometrician finds this maximum likelihood estimate is discussed briefly in the technical notes to this section.

In addition to its intuitive appeal, the maximum likelihood estimator has several desirable asymptotic properties. It is asymptotically unbiased, it is consistent, it is asymptotically efficient, it is distributed asymptotically normally, and its asymptotic variance can be found via a standard formula (the Cramer–Rao lower bound – see the technical notes to this section). Its only major theoretical drawback is that in order to calculate the MLE the econometrician must assume a *specific* (e.g., normal) distribution for the error term. Most econometricians seem willing to do this.

These properties make maximum likelihood estimation very appealing for situations in which it is impossible to find estimators with desirable small-sample properties, a situation that arises all too often in practice. In spite of this, however, until recently maximum likelihood estimation has not been popular, mainly because of high computational cost. Considerable algebraic manipulation is required before estimation, and most types of MLE problems require substantial input preparation for available computer packages. But econometricians' attitudes to MLEs have changed recently, for several reasons. Advances in computers and related software have dramatically reduced the computational burden. Many interesting estimation problems have been solved through the use of MLE techniques, rendering this approach more useful (and in the process advertising its properties more widely). And instructors have been teaching students the theoretical aspects of MLE techniques, enabling them to be more comfortable with the algebraic manipulations they require.

2.10 MONTE CARLO STUDIES

A Monte Carlo study is a simulation exercise designed to shed light on the small-sample properties of competing estimators for a given estimating problem. They are called upon whenever, for that particular problem, there exist potentially attractive estimators whose small-sample properties cannot be derived theoretically. Estimators with unknown small-sample properties are continually being proposed in the econometric literature, so Monte Carlo studies have become quite common, especially now that computer technology has made their undertaking quite cheap. This is one good reason for having a good understanding of this technique. A more important reason is that a thorough understanding of Monte Carlo studies guarantees an understanding of the repeated sample and sampling distribution concepts, which are crucial to an understanding of econometrics. Appendix A at the end of this book has more on sampling distributions and their relation to Monte Carlo studies.

The general idea behind a Monte Carlo study is to (1) model the data-generating process, (2) generate several sets of artificial data, (3) employ these data and an estimator to create several estimates, and (4) use these estimates to gauge the sampling distribution properties of that estimator. This is illustrated in figure 2.7. These four steps are described below.

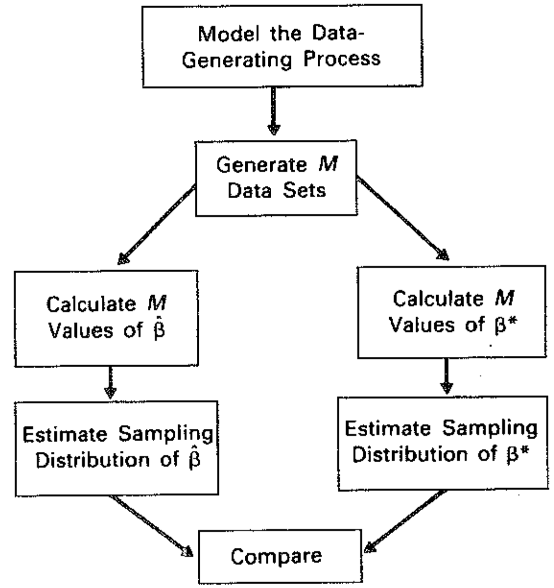

Figure 2.7 Structure of a Monte Carlo study

(1) *Model the data-generating process* Simulation of the process thought to be generating the real-world data for the problem at hand requires building a model for the computer to mimic the data-generating process, including its stochastic component(s). For example, it could be specified that N (the sample size) values of X, Z, and an error term generate N values of Y according to $Y = \beta_1 + \beta_2 X + \beta_3 Z + \varepsilon$, where the β_i are specific, known numbers, the N values of X and Z are given, exogenous, observations on explanatory variables, and the N values of ε are drawn randomly from a normal distribution with mean zero and known variance σ^2. (Computers are capable of generating such random error terms.) Any special features thought to characterize the problem at hand must be built into this model. For example, if $\beta_2 = \beta_3^{-1}$ then the values of β_2 and β_3 must be chosen such that this is the case. Or if the variance σ^2 varies from observation to observation, depending on the value of Z, then the error terms must be adjusted accordingly. An important feature of the study is that all of the (usually unknown) parameter values are *known* to the person conducting the study (because this person chooses these values).

(2) *Create sets of data* With a model of the data-generating process built into the computer, artificial data can be created. The key to doing this is the stochastic element of the data-generating process. A sample of

size N is created by obtaining N values of the stochastic variable ε and then using these values, in conjunction with the rest of the model, to generate N values of Y. This yields one complete sample of size N, namely N observations on each of Y, X, and Z, corresponding to the particular set of N error terms drawn. Note that this artificially generated set of sample data could be viewed as an *example* of real-world data that a researcher would be faced with when dealing with the kind of estimation problem this model represents. Note especially that the set of data obtained depends crucially on the particular set of error terms drawn. A different set of error terms would create a different data set *for the same problem.* Several of these examples of data sets could be created by drawing different sets of N error terms. Suppose this is done, say, 2,000 times, generating 2,000 sets of sample data, each of sample size N. These are called repeated samples.

(3) *Calculate estimates* Each of the 2,000 repeated samples can be used as data for an estimator $\hat{\beta}_3$, say, creating 2,000 estimated $\hat{\beta}_{3i}$ $(i = 1, 2, \ldots, 2{,}000)$ of the parameter β_3. These 2,000 estimates can be viewed as random "drawings" from the sampling distribution of $\hat{\beta}_3$.

(4) *Estimate sampling distribution properties* These 2,000 drawings from the sampling distribution of $\hat{\beta}_3$ can be used as data to estimate the properties of this sampling distribution. The properties of most interest are its expected value and variance, estimates of which can be used to estimate bias and mean square error.

(a) The *expected value* of the sampling distribution of $\hat{\beta}_3$ is estimated by the average of the 2,000 estimates:

$$\text{estimated expected value} = \bar{\hat{\beta}} = \left(\sum_{i=1}^{2{,}000} \hat{\beta}_{3i}\right) \Big/ 2{,}000.$$

(b) The *bias* of $\hat{\beta}_3$ is estimated by subtracting the known true value of β_3 from the average:

$$\text{estimated bias} = \bar{\hat{\beta}}_3 - \beta_3.$$

(c) The *variance* of the sampling distribution of $\hat{\beta}_3$ is estimated by using the traditional formula for estimating variance:

$$\text{estimated variance} = \sum_{i=1}^{2{,}000} (\hat{\beta}_{3i} - \bar{\hat{\beta}}_3)^2 \Big/ 1{,}999.$$

(d) The *mean square error* of $\hat{\beta}_3$ is estimated by the average of the squared differences between $\hat{\beta}_3$ and the true value of β_3:

$$\text{estimated MSE} = \sum_{i=1}^{2{,}000} (\hat{\beta}_{3i} - \beta_3)^2 \Big/ 2{,}000.$$

At stage 3 above an alternative estimator β_3^* could also have been used to calculate 2,000 estimates. If so, the properties of the sampling distribution of β_3^* could also be estimated and then compared with those of the sampling distribution of $\hat{\beta}_3$. (Here $\hat{\beta}_3$ could be, for example, the ordinary least squares estimator and β_3^* any competing estimator such as an instrumental variable estimator, the least absolute error estimator or a generalized least squares estimator. These estimators are discussed in later chapters.) On the basis of this comparison, the person conducting the Monte Carlo study may be in a position to recommend one estimator in preference to another for the sample size N. By repeating such a study for progressively greater values of N, it is possible to investigate how quickly an estimator attains its asymptotic properties.

2.11 ADDING UP

Because in most estimating situations there does not exist a "super-estimator" that is better than all other estimators on all or even most of these (or other) criteria, the ultimate choice of estimator is made by forming an "overall judgment" of the desirableness of each available estimator by combining the degree to which an estimator meets each of these criteria with a subjective (on the part of the econometrician) evaluation of the importance of each of these criteria. Sometimes an econometrician will hold a particular criterion in very high esteem and this will determine the estimator chosen (if an estimator meeting this criterion can be found). More typically, other criteria also play a role in the econometrician's choice of estimator, so that, for example, only estimators with reasonable computational cost are considered. Among these major criteria, most attention seems to be paid to the best unbiased criterion, with occasional deference to the mean square error criterion in estimating situations in which all unbiased estimators have variances that are considered too large. If estimators meeting these criteria cannot be found, as is often the case, asymptotic criteria are adopted.

A major skill of econometricians is the ability to determine estimator properties with regard to the criteria discussed in this chapter. This is done either through theoretical derivations using mathematics, part of the technical expertise of the econometrician, or through Monte Carlo studies. To derive estimator properties by either of these means, the mechanism generating the observations must be known; changing the way in which the observations are generated creates a new estimating problem, in which old estimators may have new properties and for which new estimators may have to be developed.

The OLS estimator has a special place in all this. When faced with any estimating problem, the econometric theorist usually checks the OLS estimator first, determining whether or not it has desirable properties. As seen in the next chapter, in some circumstances it does have desirable properties and is chosen as the "preferred" estimator, but in many other circumstances it does not have desirable properties and a replacement must be found. The econometrician must investigate whether the circumstances under which the OLS estimator is desirable are met, and, if not, suggest appropriate alternative estimators. (Unfortunately, in practice this is too often not done, with the OLS estimator being adopted without justification.) The next chapter explains how the econometrician orders this investigation.