EDWARD GREENBERG
CHARLES E. WEBSTER, JR.

ADVANCED ECONOMETRICS: A BRIDGE TO THE LITERATURE

WILEY SERIES IN PROBABILITY
AND MATHEMATICAL STATISTICS

ADVANCED ECONOMETRICS
A Bridge to the Literature

Edward Greenberg
Washington University, St. Louis

Charles E. Webster, Jr.
Washington University, St. Louis

John Wiley & Sons
New York *Chichester* *Brisbane* *Toronto* *Singapore*

To Joan and Lesley

Library of Congress Cataloging in Publication Data:

Greenberg, Edward, 1936–
  Advanced econometrics.

  (Wiley series in probability and mathematical statistics. Probability and mathematical statistics, ISSN 0271-6232)
  Includes bibliographical references and indexes.
  1. Econometrics.   I. Webster, Jr. Charles E. (Charles Edward), 1950–   II. Title.   III. Series.
HB139.G73   1983   330'.028   82-23770
ISBN 0-471-09077-8

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Preface

Many articles appearing in the current econometrics literature are inaccessible to students with a background of one or two semesters in mathematical statistics and an introductory course in econometrics. Our book bridges the gap between an introductory course and such journal articles. It presents background material on selected topics in mathematics and mathematical statistics and provides fairly detailed reviews of three areas of econometric research. The book is intended for courses that continue beyond an introductory econometrics course in a theoretical direction. To a great extent, it is keyed to significant articles in the field to encourage the reading of the original articles.

Part I consists of two chapters that present probabilistic concepts of convergence, maximum likelihood estimation, asymptotic expansions of probability functions, and methods of approximating moments. These widely used large sample and approximation techniques are not adequately covered in most introductory econometrics texts. We present relationships between plims, asymptotic moments, and moments of limiting distributions. An introduction to Edgeworth approximations to distribution and density functions is included, along with a discussion of methods of approximating moments. Although the material on convergence and maximum likelihood estimation is primarily a review of topics covered in mathematical statistics and econometrics courses, the discussion of approximation techniques is likely to be the student's first exposure to these topics. In view of their widespread use, particularly in connection with distributions of estimators of parameters in simultaneous equation models, it is important that students know about them. Appendixes on order of magnitude (o and O), complex variables, and characteristic functions provide necessary background.

Part II consists of three chapters on the analysis of time series models. The first two of these present the theory and estimation methods for the ARIMA models popularized by Box and Jenkins. Numerous graphs depict the typical pattern of data generated by low order autoregressive and moving average models and the autocovariance and partial autocovariance functions of these models. The fundamental theoretical concepts of stationarity and invertibility are explained both intuitively and formally. The third chapter of this part takes up the Zellner-Palm synthesis of time series and standard econometric simultaneous equation models, and an illustration of the approach is provided. The chapter concludes with an explanation and a critique of causality tests of the type proposed by Granger and Sims.

Part III is concerned with estimation, particularly of regression coefficients, under a mean squared error (quadratic loss function) criterion. After presenting background material on decision theory concepts and the noncentral χ^2 distribution, we turn to the mean squared error performance of a number of estimators: ordinary least squares, restricted least squares, James-Stein and variants, preliminary test, double k-class, and ridge regression. The emphasis is on analytical results and comparisons of the estimators' performances. A discussion of the use of these estimators in econometric research is included.

Part IV examines the estimation of parameters of simultaneous equation models. It emphasizes topics that are not covered in most introductory textbooks because they require considerably more mathematical and statistical technique than most econometric students possess. Appendixes on the ratio of normal variates, the noncentral Wishart distribution, and hypergeometric and confluent hypergeometric functions provide background material. The first chapter deals with identification and contains results that generalize the one equation, linear restriction case that is included in most introductory texts. The next chapter considers the special case of two included endogenous variables. It derives in detail the exact distribution of ordinary least squares and two-stage least squares estimators and the moments of k-class estimators. In addition, references are provided for results with other estimators, and approximation methods are applied to derive relatively tractable expressions for distributions and moments. The third chapter presents in detail a theorem on existence of moments of k-class estimators with an arbitrary number of included endogenous variables. It also includes a summary of and reference to other topics found in the current literature.

We are happy to acknowledge Professor Arthur Denzau and the help of several of our students, who made comments at various stages: Alvin Hsieh, Dosung Chung, Mark Stover, Blair Harcourt, and Howard Neiman. We also thank Professors R. Carter Hill, Albert Link, Malcolm Dowling, and Jan Kmenta for their very useful comments on the project in its early stages, and Professor Jerry Thursby who made many detailed comments on our completed draft.

Typing was expertly and (usually) cheerfully done by Mary Kastens, Patricia Curtis, Anne Schroeder, and Candace Schap.

Edward Greenberg
Charles E. Webster, Jr.

LIST OF ABBREVIATIONS, FUNCTIONS, AND SYMBOLS

Abbreviations

a.b.p.	Asymptotically bounded in probability
ACF	Autocorrelation function
AR	Autoregressive
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
c.f.	Characteristic function
c.r.	Characteristic root
c.v.	Characteristic vector
CLT	Central limit theorem
CRLB	Cramer-Rao lower bound
d.f.	Distribution function
i.i.d.	Identically, independently distributed
IV	Instrumental variables
LIML	Limited information maximum likelihood
MA	Moving average
m.g.f.	Moment generating function
MLE	Maximum likelihood estimator
$N'_\epsilon(x)$	Deleted neighborhood of x; $\{u: \|u - x\| < \epsilon$ and $u \neq x\}$
$o(b_t)$	"$a_t = o(b_t)$" indicates a_t is of *smaller order* than b_t
$O(b_t)$	"$a_t = O(b_t)$" indicates a_t is *at most of order* b_t
$o_p(a_t)$	"$Y_t = o_p(a_t)$" indicates Y_t is of smaller order in probability than a_t
$O_p(a_t)$	"$Y_t = O_p(a_t)$" indicates Y_t is at most of order in probability as a_t
OLS	Ordinary least squares
PACF	Partial autocorrelation function
$P[A]$	Probability of the event A
plim	Probability limit
$r(A)$	Rank of matrix A
r.v.	Random variable

SLLN	Strong law of large numbers
3SLS	Three-stage least squares
2SLS	Two-stage least squares
WLLN	Weak law of large numbers

Functions

$\arg Z$	Argument of the complex number Z
$B(a, b)$	Beta function
χ_k^2	Central χ^2 distribution with k degrees of freedom
$\chi'^2(n, \lambda)$	Noncentral χ^2 distribution with n degrees of freedom and noncentrality parameter λ
$\mathrm{Cov}(\mathbf{X})$	Covariance matrix of the random vector $\mathbf{X}$
$E(X)$	Expected value of X (scalar or vector)
$F(k, n)$	Central F-distribution with k and n degrees of freedom
${}_1F_1(a, b; z)$	Confluent hypergeometric function
${}_2F_1(a, b; c; z)$	Hypergeometric function
$\Gamma(a)$	Gamma function
$I(\boldsymbol{\theta})$	Information matrix
$\mathrm{Im}\, Z$	Imaginary part of the complex number Z
$l(\boldsymbol{\theta}; \mathbf{x})$	Log-likelihood function
$\hat{l}(\boldsymbol{\theta}; \mathbf{x})$	Log-likelihood function evaluated at the maximum likelihood estimator
$L(\boldsymbol{\theta}; \mathbf{x})$	Likelihood function
$\hat{L}(\boldsymbol{\theta}; \mathbf{x})$	Likelihood function evaluated at the maximum likelihood estimator
$N(\mu, \sigma^2)$	Normal distribution with expected value μ and variance σ^2
$N_p(\boldsymbol{\mu}, \Sigma)$	p-dimensional multivariate normal distribution with expected value $\boldsymbol{\mu}$ and covariance matrix Σ
$\mathrm{Re}\, Z$	Real part of the complex number Z
$\mathrm{row}(V, W)$	The $1 \times mn$ vector formed from the $m \times p$ matrix V and the $p \times n$ matrix W as $(\mathbf{V}_1W, \mathbf{V}_2W, \ldots, \mathbf{V}_mW)$, where $\mathbf{V}_i$ is the ith row of V
$\mathrm{Var}(X)$	Variance of the random variable X
$\mathrm{vec}(A)$	The $1 \times mn$ vector formed from the $m \times n$ matrix A as $(\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m)$, where $\mathbf{a}_i$ is the ith row of A
$W_p(\Sigma, N)$	p-dimensional Wishart distribution with scale matrix Σ and N degrees of freedom
$W_p'(\Sigma, N, T)$	p-dimensional noncentral Wishart distribution with scale matrix Σ, means sigma matrix T and N degrees of freedom

Symbols

$\sim$	i. "$x \sim f(x)$" indicates that the random variable x (scalar or vector) has the density function $f(x)$.
	ii. "$h(x) \sim \Sigma a_n \phi_n(x),\ x \to c$" indicates that $\Sigma a_n \phi_n(x)$ is an asymptotic expansion for $h(x)$ as $x \to c$
$\xrightarrow{\text{a.s.}}$	Convergence almost surely
$\xrightarrow{\text{d}}$	Convergence in distribution
$\xrightarrow{\text{p}}$	Convergence in probability
$\xrightarrow{\text{q.m.}}$	Convergence in quadratic mean
$\lvert Z \rvert$	Modulus of the complex number Z
$\lVert \mathbf{X} \rVert$	Euclidean norm of vector $\mathbf{X}$; $(\mathbf{X}'\mathbf{X})^{1/2}$
$(a)_n$	$\dfrac{\Gamma(a+n)}{\Gamma(a)} = (a+n-1)(a+n-2)\ldots a$

Contents

PART I

Topics in Large Sample Theory and Approximation Methods

Chapter 1 is primarily a review of convergence concepts used in probability theory and of the maximum likelihood method of estimation. Chapter 2 illustrates a number of methods that are used to approximate probability distributions and moments of functions of random variables.

Although only a few of those topics are extensively used in the remainder of the book, a basic knowledge of the topics covered in Part I is necessary for understanding and contributing to current research in econometrics.

CHAPTER 1

A Review of Asymptotic Distribution Theory and Maximum Likelihood Estimation

We begin our study of asymptotic distribution theory with an examination of sequences of random variables. Several features of such sequences may be studied: Are there concepts of convergence which might be applied to random variables (r.v.'s)? Can anything be said about the behavior of the sequence of corresponding distribution functions, $\{F_T(x)\}$? Can anything be said about the sequence of moments, $\{E(X_t^r)\}$? Can approximations be found to complicated distribution functions as $T \to \infty$? Precise definitions for these concepts and the relationships between them have been extensively studied.

Relevance for econometricians arises from the fact that X_T may often be interpreted as an estimator for some parameter based on a sample size of T, and the behavior of the sequence for large T may be regarded as the behavior of the estimator for large sample sizes. Since it is often easier to investigate the properties of X_T for large T than to ascertain exact (or small sample) distributions, asymptotic concepts have been extensively utilized in econometrics. The questions of interest are whether X_T has any desirable statistical properties as T increases and, for purposes of establishing confidence intervals or testing hypotheses, what is the distribution of X_T for large T.

In the sections that follow we first examine various convergence concepts, the relationships between the convergence concepts, and the conditions necessary for a sequence of random variables to converge. Using these concepts of convergence we then present a number of central limit theorems showing the conditions under which a sequence of random variable converges to a random variable that is normally distributed. In the final section of this chapter we examine the properties of maximum likelihood estimators for large sample size.[1]

1.1 Convergence Concepts

Since several of the convergence concepts discussed in detail below are bewilderingly similar, some general introductory comments on probability theory may be useful in understanding the differences between them.[2] We begin by defining Ω as the set of all possible outcomes of an experiment, and ω is used to denote a particular outcome. A

[1]Econometrics texts that treat these topics in some detail are Dhrymes (1970, Chapter 3) and Theil (1971, Chapter 8).

[2]Pfeiffer (1978) is an excellent introduction to the basic probability theory used in this book. It contains the necessary restrictions on the structure of subsets of Ω and on the functions that can be r.v.'s.

random variable, $X(\omega)$, is a real valued function with domain Ω that assigns a numerical value to each outcome. In statistical applications, Ω is the sample space, $\omega \in \Omega$ is a possible outcome, and $X(\omega)$ is a numerical value assigned to outcome ω.

As an example, consider the experiment of repeatedly tossing a coin, where the possible outcomes are a head (H) or a tail (T). Then each ω is an infinite sequence of H's and T's, and Ω is the set of all such sequences. An example of a random variable, $X(\omega)$, is the proportion of heads in the first ten tosses, which is computed from the number of H's in the first ten positions of ω.

More than one function (each of which is a r.v.) may be defined for a value of ω. A *sequence of random variables* is a sequence of real valued functions, $X_1(\omega), \ldots,$ $X_T(\omega), \ldots$, each of which has the same ω as its argument. For example, returning to the coin-tossing experiment, $X_1(\omega)$ might be the proportion of heads in one toss, $X_2(\omega)$ the proportion in two tosses, and $X_T(\omega)$ the proportion in T tosses. Thus, if $\omega = \{H, T, T, T, H, H, \ldots\}$, then $X_1(\omega) = 1.00$, $X_2(\omega) = .50$, $X_3(\omega) = .33$, $X_4(\omega) = .25$, $X_5(\omega) = .40$, and so on. A second example would be to define ω as a particular outcome of gross national product (GNP) now, next year, and into the future. A particular ω, say ω^*, may refer to a time path of GNP of 85 in period 1, 95 in period 2, 108 in period 3, and so on. Ω would then refer to all possible time paths of GNP from now into the future. We could then define a sequence of random variables $X_i(\omega)$ where X_i is the average value of GNP after i periods. Hence, for our particular series we would have $X_1(\omega^*) = 85$, $X_2(\omega^*) = 90$, $X_3(\omega^*) = 96$, and so on.

Because it is often of interest to see whether a sequence of random variables converges, we next discuss several senses in which the sequence $\{X_T\}$ may be said to converge.

1.1.1 Convergence Almost Surely (a.s.)[3]

Before presenting the definition of "convergence almost surely" it is necessary to explain some notation. Consider the sequence of functions, $X_1(\omega), X_2(\omega), \ldots$, where each $X_i(\omega)$ is a random variable evaluated at the point ω. It may happen that the sequence $\{X_T(\omega)\}$ converges to the point $X(\omega)$ in the sense that, for sufficiently large T, the difference $|X_T(\omega) - X(\omega)|$ may be made arbitrarily small and remains arbitrarily small (less than ϵ) as T increases. That is, $|X_{T+i}(\omega) - X(\omega)| \leq \epsilon$ for $i = 0, 1, \ldots$ and $\epsilon > 0$. If that occurs, we write $\lim_{T\to\infty} X_T(\omega) = X(\omega)$. This notation is short for the statement: for every $\epsilon > 0$, there exists a $T_0(\epsilon)$ such that

$$|X_T(\omega) - X(\omega)| \leq \epsilon, \qquad \text{whenever} \quad T > T_0(\epsilon).$$

Now consider the set, A, of ω for which the convergence takes place,

$$A = \{\omega: \lim_{T\to\infty} X_T(\omega) = X(\omega)\}.$$

The probability of the occurrence of A, $P(A)$, is often expressed as

$$P(A) = P[\lim_{T\to\infty} |X_T(\omega) - X(\omega)| \leq \epsilon].$$

The first concept of convergence is concerned with $P(A)$:

[3]This type of convergence is also called "convergence with probability one."

Definition

$X_T(\omega)$ *converges almost surely to* X if $P[\lim_{T\to\infty} |X_T(\omega) - X(\omega)| \leq \epsilon] = 1$, for every $\epsilon > 0$. This will be denoted by $X_T(\omega) \xrightarrow{\text{a.s.}} X$. $X_T(\omega)$ and $X(\omega)$ may be scalars or column vectors; in the latter case, $\mathbf{X}_T(\omega)$ and $\mathbf{X}(\omega)$ will be of the same dimension and $|\mathbf{X}_T(\omega) - \mathbf{X}(\omega)|$ should be interpreted as the Euclidean norm, $[\mathbf{X}_T(\omega) - \mathbf{X}(\omega)]'[\mathbf{X}_T(\omega) - \mathbf{X}(\omega)]^{1/2}$. Moreover, X may be either a constant (or vector of constants) or a r.v. (or vector of r.v.'s).

An equivalent definition of convergence a.s. is helpful in understanding the concept. To develop this, consider the set of ω for which convergence *fails* to take place. $X_T(\omega)$ does not converge to $X(\omega)$ if for some $\epsilon > 0$, $|X_{T+i}(\omega) - X(\omega)| > \epsilon$ for at least one i, no matter how large T is. That is, either $|X_{T+1}(\omega) - X(\omega)| > \epsilon$ or $|X_{T+2}(\omega) - X(\omega)| > \epsilon$ or . . . for every T. For convergence a.s., the set of such ω has probability zero.

An example of convergence a.s. is the "strong law of large numbers" (SLLN): Let $\{Y_i\}$ be a sequence of r.v.'s and $\{X_T\}$ a sequence of partial sums of the Y_i; that is, $X_T = 1/T \sum_{i=1}^{T} Y_i$. In addition, assume that $X = \lim_{T\to\infty} E(X_T)$ exists and is finite. If $X_T(\omega) \xrightarrow{\text{a.s.}} X$, the sequence $\{Y_i\}$ is said to obey the SLLN. Two theorems that give sufficient conditions for a sequence of r.v.'s to obey SLLN are next quoted:

Theorem 1.1.1 (Kolmogoroff)

If a sequence of mutually independent variables $\{Y_i\}$ satisfies

$$\sum_{n=1}^{\infty} \frac{\text{Var}(Y_n)}{n^2} < \infty,$$

then it obeys the SLLN.

Proof

See Gnedenko (1973, p. 215). □

Theorem 1.1.2

The existence of $E(Y_i)$ is a necessary and sufficient condition for applying the SLLN to $\{Y_i\}$, where $\{Y_i\}$ is a sequence of identically distributed and mutually independent r.v.'s.

Proof

See Gnedenko (1973, p. 216). □

Another example of convergence a.s. is provided by the following: Let $X \sim N(0, 1)$ and

$$X_T = X + \frac{1}{T}.$$

Since $X_T - X = 1/T$, it is clear that $|X_T(\omega) - X(\omega)| \leq \epsilon$ for every T such that $1/T \leq \epsilon$. Therefore, the set $\{\omega: \lim|X_T(\omega) - X(\omega)| > \epsilon\}$ is empty. Since the empty set has probability zero, we have $X_T \xrightarrow{\text{a.s.}} X$.

At the risk of belaboring the point, assume ω is an outcome of an experiment which generates a value from a $N(0, 1)$ variable, $X(\omega)$. If $X_T(\omega) = X(\omega) + 1/T$, a particular outcome, ω, determines $X(\omega)$, and then $X_T(\omega)$ is determined simply by adding $1/T$. Therefore, as T becomes large, $|X_T(\omega) - X(\omega)|$ becomes very small: $X_T(\omega)$ and $X(\omega)$ are virtually the same random variable for large enough T; thus $X_T(\omega) \xrightarrow{\text{a.s.}} X(\omega)$, where the latter is a r.v.

Convergence a.s. is a strong convergence concept—any sequence of r.v.'s that converges a.s. also converges "in probability" and "in distribution," two types of convergence discussed below. Since convergence a.s. is a strong property, rather stringent sufficient conditions must be placed on the r.v. to achieve it. Perhaps to avoid such restrictive assumptions, econometricians have preferred to work with the weaker types of convergence discussed in the following sections.

1.1.2 Convergence in Probability

Definition

The sequence of r.v.'s $\{X_T\}$ *converges in probability to zero* if for all $\epsilon > 0$, $\lim_{T\to\infty} P[|X_T| > \epsilon] = 0$. X_T may be a scalar or a vector; if the latter, the convergence is to the zero vector of suitable dimension. This type of convergence will be denoted by $X_T \xrightarrow{\text{P}} 0$. An equivalent definition is that X_T converges in probability to zero if for all $\epsilon > 0$ and $\delta > 0$, $P[|X_T| \leq \epsilon] \geq 1 - \delta$ for sufficiently large T.

For this type of convergence we consider the set of points $\{\omega: |X_T(\omega)| > \epsilon\}$ and its probability. To satisfy the definition there must exist a value of T, say T_0, such that for $T > T_0$, $P[\{\omega: |X_T(\omega)| > \epsilon\}]$ is arbitrarily small. Convergence in probability implies that for $T > T_0$, *each* of the sets of ω (events), $\{\omega: |X_0| > \epsilon\}$, $\{\omega: |X_{0+1}| > \epsilon\}$, . . . , have arbitrarily small probabilities, but not that the probability of the *union* of these events is arbitrarily small. The latter, as we have seen, is the definition of convergence a.s. To see the relationship between convergence a.s. and convergence in probability, define $S_{i,\epsilon} = \{\omega: |X_{0+i}(\omega) > \epsilon\}$ and assume that $X_T \xrightarrow{\text{a.s.}} 0$. Then there exists a T_0 such that $P[S_{0,\epsilon} \cup S_{1,\epsilon} \cup \cdots S_{N,\epsilon} \cup \cdots] < \delta$. But since the probability of the union of a set of events is greater than or equal to the probability of any of the individual events, we must also have $P[S_{i,\epsilon}] < \delta$; hence there exists a T_0 that makes $P[|X_T(\omega)| > \epsilon] < \delta$. This proves the following theorem.

Theorem 1.1.3

Convergence a.s. implies convergence in probability.

We next provide an example showing that the converse is not true. Although the example is rather difficult, it should be studied to appreciate the difference between the two convergence concepts. This example is a probabilistic interpretation of a standard