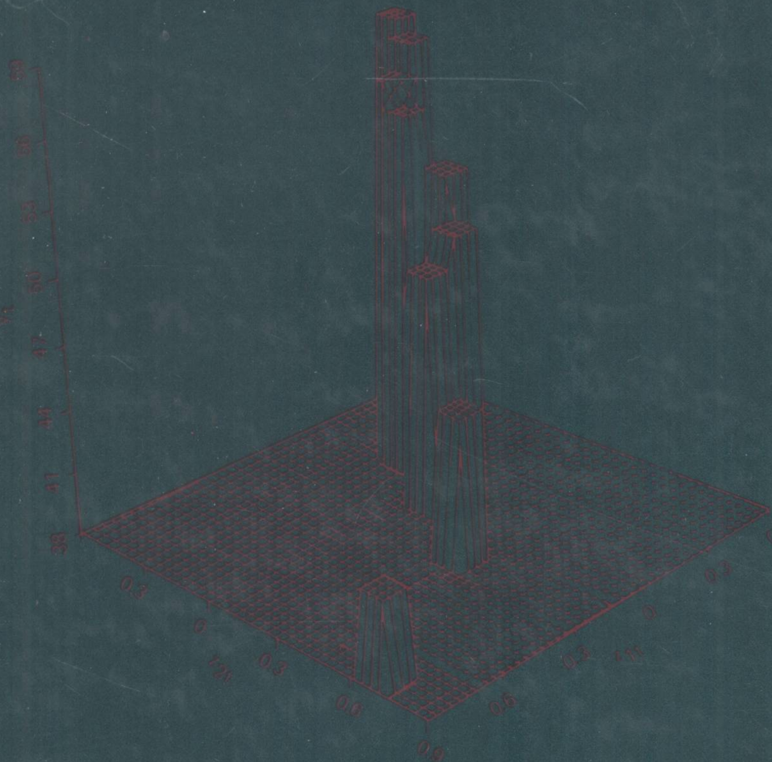
STATISTICS: textbooks and monographs volume 41



RECENT ADVANCES IN REGRESSION METHODS

HRISHIKESH D. VINOD
AMAN ULLAH

RECENT ADVANCES IN REGRESSION METHODS

Hrishikesh D. Vinod

Economic Analysis Section
American Telephone and Telegraph Company
Piscataway, New Jersey

Aman Ullah

Department of Economics
University of Western Ontario
London, Ontario, Canada

MARCEL DEKKER, INC. New York and Basel

Library of Congress Cataloging in Publication Data

Vinod, Hrishikesh D., [date]
Recent advances in regression methods.

(Statistics, textbooks and monographs ; v. 41)
Includes index.
1. Regression analysis. I. Ullah, Aman. II. Title.
III. Series.
QA278.2.V56 519.5'36 81-15079
ISBN 0-8247-1574-8 AACR2

MARCEL DEKKER, INC.

270 Madison Avenue, New York, New York 10016

Current printing (last digit):

10 9 8 7 6 5 4 3 2 1

PRINTED IN THE UNITED STATES OF AMERICA

RECENT ADVANCES IN REGRESSION METHODS

STATISTICS: Textbooks and Monographs

A SERIES EDITED BY

Volume 1: The Generalized Jackknife Statistic, *H. L. Gray and W. R. Schucany*

Volume 2: Multivariate Analysis, *Anant M. Kshirsagar*

Volume 3: Statistics and Society, *Walter T. Federer*

Volume 4: Multivariate Analysis: A Selected and Abstracted Bibliography, 1957-1972, *Kocherlakota Subrahmaniam and Kathleen Subrahmaniam* (out of print)

Volume 5: Design of Experiments: A Realistic Approach, *Virgil L. Anderson and Robert A. McLean*

Volume 6: Statistical and Mathematical Aspects of Pollution Problems, *John W. Pratt*

Volume 7: Introduction to Probability and Statistics (in two parts) Part I: Probability; Part II: Statistics, *Narayan C. Giri*

Volume 8: Statistical Theory of the Analysis of Experimental Designs, *J. Ogawa*

Volume 9: Statistical Techniques in Simulation (in two parts), *Jack P. C. Kleijnen*

Volume 10: Data Quality Control and Editing, *Joseph I. Naus*

Volume 11: Cost of Living Index Numbers: Practice, Precision, and Theory, *Kali S. Banerjee*

Volume 12: Weighing Designs: For Chemistry, Medicine, Economics, Operations Research, Statistics, *Kali S. Banerjee*

Volume 13: The Search for Oil: Some Statistical Methods and Techniques, *edited by D. B. Owen*

Volume 14: Sample Size Choice: Charts for Experiments with Linear Models, *Robert E. Odeh and Martin Fox*

Volume 15: Statistical Methods for Engineers and Scientists, *Robert M. Bethea, Benjamin S. Duran, and Thomas L. Boullion*

Volume 16: Statistical Quality Control Methods, *Irving W. Burr*

Volume 17: On the History of Statistics and Probability, *edited by D. B. Owen*

Volume 18: Econometrics, *Peter Schmidt*

Volume 19: Sufficient Statistics: Selected Contributions, *Vasant S. Huzurbazar (edited by Anant M. Kshirsagar)*

Volume 20: Handbook of Statistical Distributions, *Jagdish K. Patel, C. H. Kapadia, and D. B. Owen*

Volume 21: Case Studies in Sample Design, *A. C. Rosander*

Volume 22: Pocket Book of Statistical Tables, *compiled by R. E. Odeh, D. B. Owen, Z. W. Birnbaum, and L. Fisher*

Volume 23: The Information in Contingency Tables, *D. V. Gokhale and Solomon Kullback*

Volume 24: Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods, *Lee J. Bain*

Volume 25: Elementary Statistical Quality Control, *Irving W. Burr*

Volume 26: An Introduction to Probability and Statistics Using BASIC, *Richard A. Groeneveld*

Volume 27: Basic Applied Statistics, *B. L. Raktoe and J. J. Hubert*

Volume 28: A Primer in Probability, *Kathleen Subrahmaniam*

Volume 29: Random Processes: A First Look, *R. Syski*

Volume 30: Regression Methods: A Tool for Data Analysis, *Rudolf J. Freund and Paul D. Minton*

Volume 31: Randomization Tests, *Eugene S. Edgington*

Volume 32: Tables for Normal Tolerance Limits, Sampling Plans, and Screening, *Robert E. Odeh and D. B. Owen*

Volume 33: Statistical Computing, *William J. Kennedy, Jr. and James E. Gentle*

Volume 34: Regression Analysis and Its Application: A Data-Oriented Approach, *Richard F. Gunst and Robert L. Mason*

Volume 35: Scientific Strategies to Save Your Life, *I. D. J. Bross*

Volume 36: Statistics in the Pharmaceutical Industry, *edited by C. Ralph Buncher and Jia-Yeong Tsay*

Volume 37: Sampling from a Finite Population, *J. Hájek*

Volume 38: Statistical Modeling Techniques, *S. S. Shapiro*

Volume 39: Statistical Theory and Inference in Research, *T. A. Bancroft and C.-P. Han*

Volume 40: Handbook of the Normal Distribution, *Jagdish K. Patel and Campbell B. Read* (in press)

Volume 41: Recent Advances in Regression Methods, *Hrishikesh D. Vinod and Aman Ullah*

OTHER VOLUMES IN PREPARATION

To my parents, Appa and Vahini,
and my wife, Arundhati (H.D.V.)

To my wife, Shobha (A.U.)

PREFACE

Since the exposition of the Gauss Markov theorem, and especially during the past fifty years or so, practically all fields in the natural and social sciences have widely and sometimes blindly used the Ordinary Least Squares (OLS) method. It gives us the Best Linear Unbiased Estimator (BLUE), which is also equivalent to the maximum likelihood (ML) estimator for estimation of the underlying normal regression relationship. It was in the late 1950's that Charles Stein brought to the attention of statisticians the fact that there is a better alternative to OLS under certain conditions. For a decade, his surprising result went largely unnoticed. In the late 1960's a series of papers was published giving interpretations and variants of what began to be called the Stein-Rule estimators. All these developments were primarily aimed at suggesting a family of biased estimators with potentially smaller Mean Squared Error (MSE) than OLS. Independent development parallel to this came from Hoerl and Kennard in 1970 who suggested another biased ridge regression estimator with potentially smaller MSE as an alternative to OLS. Ridge regression is often thought to be particularly applicable when the applied researchers in any field of natural or social science face the problem of multicollinearity in their data — a serious disease. Again, a series of papers appeared on the scene soon after the seminal work on ridge regression by Hoerl and Kennard was published.

Thus, "improved estimation" of the linear regression model has been a focus of scholarly writing over the past twenty-five years, and particularly in the past decade. The number of contributions has been so numerous that almost anyone who is not an active researcher in this area is finding the pace too fast to keep up with. Graduate students are also discovering that more and more of their required readings in standard courses in Statistics, Econometrics, Biomedicine, Psychology, Engineering, and so on are covering these topics; as a rule they have to read them from various journals. Often, the local libraries do not even subscribe to the journals from otherwise remote disciplines that may contain the appropriate references. An increasing number of university departments in various sciences are offering courses involving applications of Stein-Rule, ridge regression, etc., on regression models with their own data.

This book attempts to bring together the recent developments in the area of "improved estimation" and inference in linear models in the form of a graduate level textbook or a supplementary text. We hope that this book will be an important step in bridging the above mentioned gap in the published literature.

Our intended readers are graduate students in Statistics, Economics, Psychology, Bio-medicine, and Engineering, among others. The book can serve as a text for theoretical and applied courses in Statistics and Econometrics. We hope that this book will also serve as a handy guide to these somewhat difficult theoretical advances for the non-experts who are merely interested in an overview of the theory with an emphasis on potential applications, or as a useful starting point for a more serious research adventure. For the experts in the area it is intended to be a convenient reference book.

The book has thirteen chapters, most of which are based upon the published papers in numerous learned journals. Some of these papers were written by the authors themselves. The chapters highlight the important aspects of the main contributions, comment on the results obtained, and establish a relationship with earlier results. We include the technical aspects of the material in the simplest possible way, and discuss applications. This is one of the reasons why we think that this book is somewhat different from many other advanced books. We give a unified treatment of theory and tests by using a restricted least squares model. The shrinkage factors are used to compare various shrinkage type, ridge and Stein-Rule estimators. There are a number of sections, explanatory notes, figures, etc. included in this book which are simply unavailable in other books. For example, the canonical reduction of the original regression model using the singular value decomposition is explained, which is rarely mentioned in the available textbooks.

The discussion of multicollinearity in the usual textbooks is often misleading and incomplete. The use of eigenvectors and eigenvalues of the correlation matrix among regressors is more reliable than certain other *ad hoc* measures of multicollinearity. This book explains the related material to practitioners in simple geometrical terms.

The Stein-Rule estimator for the mean of the normal variable and its relevance to ordinary regression is not clear from the published literature. Stein's "unbiased" estimate of the Mean Squared Error of arbitrary biased estimators has considerable practical relevance, which is clarified in this book.

In later chapters the biased estimation techniques based on Stein-Rule or ridge regression is shown to lead to "improvements" over the usual estimators in the so-called distributed lag models, the models with autocorrelation and heteroscedasticity, Zellner's seemingly unrelated regression (SUR) equations and simultaneous equations models.

When there are two or more dependent variables, Hotelling's canonical correlation analysis represents a natural multivariate extension of the regression model. This is extensively used in Psychology, Education, etc. However, the coefficients of the fitted model are known to be highly unstable with respect to data perturbations. Some of these difficulties can be avoided by using ridge regression type ideas. This is also true for the so-called discriminant analysis. Our twelfth chapter shows that ridge regression type modifications can make such multivariate techniques more meaningful to the practitioner.

Our final chapter deals with non-normal errors where we have included several new minimax-type results, and a short discussion of a few aspects of robust regression methods.

This book makes extensive use of Kadane's small-sigma asymptotics to analyze the sampling properties of various newer estimators. We have included an explanatory note in Chapter 6 to give an elementary discussion of this topic. We are encouraged by the fact that the main results based on small sigma asymptotics are identical with the corresponding exact result. We recommend it as a valuable research tool. We have also made considerable use of Bayesian methods including Lindley's hyperparameter model. We offer an integration of classical and Bayesian methods without being doctrinaire.

ACKNOWLEDGMENTS

We would like to thank L. Magee, A. Hoque of University of Western Ontario for reading entire early drafts and M. Leques of AT&T for help in preparing the early chapters of the manuscript. R. L. Obenchain, W. E. Taylor, and P. S. Brandon of Bell Labs and M. Burnside and P. B. Linhart of AT&T, have each commented on at least one chapter. C. L. Mallows of Bell Laboratories commented on about half the book. V. K. Srivastava from University of Lucknow, India, has commented in detail on some sections of the manuscript. We are grateful to Chandar, Farebrother, Obenchain and others for permitting us to use some of their unpublished material. David Hendry made critical comments of Chapter 5. R. A. L. Carter commented on some early chapters. M. Gower, G. Moore, C. Patuto and E. Quinzel typed parts of the manuscript with patience, efficiency and skill.

Vinod wishes to thank Bell Laboratories and AT&T for research support. Ullah acknowledges the research support of NSERC, and a leave fellowship from Canada Council. A part of his work on the book was done during his visit to Monash University, Australia and C.O.R.E. in Belgium. The research facilities at both these places are greatly appreciated.

Although many colleagues, students and friends have contributed directly or indirectly in the preparation of this book, they bear no responsibility whatever for any shortcomings or errors.

Hrishikesh D. Vinod
Aman Ullah

TABLE OF CONTENTS

1

Linear Regression Model

1.1 INTRODUCTION

The least squares regression model is used for studying various types of relations including technical, behavioral, static, and dynamic ones, both at the individual (micro) and collective (macro) levels. Adrian Marie Legendre [1805] was the first to publish results concerning the method of least squares, although others may have used it before him; e.g. Galileo Galilei [1632] came very close to proposing a theory of errors related to the least squares method. Gauss [1806] postulates that when any number of equally good data regarding an unknown quantity are available, their arithmetic mean is the "most probable" value. From this, Gauss derives his normal law of error or what we call the Gaussian or the normal distribution. Later Gauss [1823] and Markoff [1900] developed the theory of least squares for estimation of parameters in a general linear regression model which has proved to be useful in a great many fields of application.

In this chapter we study the general linear regression between a dependent variable and a set of explanatory variables (regressors). Various canonical forms of this relationship are given. Further, several concepts related to canonical reductions are explained both algebraically and geometrically by using a simple example of two regressors. Since this topic is covered extensively in various textbooks, we have concentrated on items that may not be generally available.

1.2 SPECIFICATION OF THE MODEL

Consider a linear regression model as

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \cdots + \beta_p x_{pt} + u_t \tag{1}$$

where $\beta_1, \ldots, \beta_p$ are unknown regression coefficients; y_t are observations on the dependent variable; x_{it} are observations on the p regressors; u_t are true unknown errors (shocks or disturbances); $t = 1, \ldots, T$ is the index for numbering observations, and $i = 1, \ldots, p$ is the index for numbering the variables and their coefficients. We may rewrite (1) in matrix notation as

$$y = X\beta + u \tag{2}$$

where y is a $T \times 1$ vector of observable random variables; X is a $T \times p$ matrix of known constants (non-stochastic regressors); β is a $p \times 1$ vector of unknown regression coefficients; u is a $T \times 1$ vector of disturbances.

We often use the following conventional assumptions:

Assumption 1 Non-stochastic X

X is a non-stochastic matrix of regressors.

Assumption 2 Full rank of $X'X$

The rank of X is p.

Assumption 3 Asymptotically full rank of $X'X/T$

$$\lim_{T \to \infty} \frac{X'X}{T} = Q$$

where Q is a finite and non-singular matrix of rank p. This assumption is required primarily for consistency and other asymptotic results in Econometrics mentioned in later chapters.

Assumption 4 Normality of the disturbances

$$u \sim N(0, \sigma^2 I).$$

The $T \times 1$ vector u of errors has a multivariate normal distribution.

This assumption is required primarily for various tests of significance. It also implies that the disturbances are homoscedastic, i.e. $V(u_t) = \sigma^2$ for all $t = 1, \ldots, T$, and that they are serially independent, i.e. $cov(u_t, u_{t'}) = 0$ for all $t \neq t' = 1, \ldots, T$.

Most of the estimators discussed in the book have been given non-Bayesian interpretations and do not need the normality assumption for studying their properties. Chapter 13 deals with the implications of non-normality. This assumption of normality is generally essential for maximum likelihood estimation. Explanatory Note 1.1 at the end of this chapter clarifies the concept of maximum likelihood estimation for readers who might not be familiar with it.

Under an appropriate subset of these assumptions the conditional expectation of y given the regressors X, and the conditional variance covariance matrix of y are:

$$E(y|X) = X\beta \; ; \; V(y|X) = E[y - E(y|X)][y - E(y|X)]' = \sigma^2 I \qquad (3)$$

Thus we have $p+1$ "true unknown" parameters β and σ^2 in our model (2).

A condition that will be used occasionally is that one of the regressors is a constant term; that is, $x_{it} = 1$ for all t and for some $1 \leq i \leq p$. Alternatively, we could add an intercept in (2) and write $y = \beta_0 + X\beta + u$. This has been taken up in a later section.

The ordinary least squares (OLS) estimator of $\beta_1, \ldots, \beta_p$ which minimizes

$$S = \sum_{t=1}^{T} (y_t - \beta_1 x_{1t} - \cdots - \beta_p x_{pt})^2 \, , \qquad (4)$$

or in matrix notation: $S = (y - X\beta)'(y - X\beta)$, is obtained by taking the derivative of S with respect to β and equating it to zero. This gives for $\beta = b$,

$$X'Xb = X'y \qquad (5)$$

a set of equations known as the least squares normal equations. Now, if X is of rank p, $(X'X)^{-1}$ exists and we have

$$b = (X'X)^{-1}X'y \qquad (6)$$

which is called the OLS estimator. This value of β corresponds to a minimum of S because the second-order partial derivative of S with respect to β is a positive definite matrix $2X'X$.

Once β has been estimated by b, we can write

$$\hat{u} = y - Xb \qquad (7)$$

as the estimator for the error (residual) vector u. Further, the sum of squares of estimated residuals divided by $T - p$, viz.,

$$s^2 = \frac{\hat{u}'\hat{u}}{T - p} \qquad (8)$$

can be shown to be a consistent and unbiased estimator of σ^2. We note that the sum of the estimated residuals (unless one of the regressors is a constant) is not zero. However, $X'\hat{u} = 0$. Thus, on the basis of the estimated regression $y = Xb + \hat{u}$,

$$y'y = b'X'Xb + \hat{u}'\hat{u} \qquad (9)$$

where $y'y$ is the total sum of squares (SST), $b'X'Xb$ is the sum of squares due to regression (SSR), and $\hat{u}'\hat{u}$ is the sum of squares due to errors (SSE). The multiple correlation coefficient (a measure of goodness-of-fit) is then defined as